```
                      GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   June 10, 2000, 07:35:42 ; Search time 29.96 Seconds
                                     (without alignments)
                                     175.511 Million cell updates/sec

Title:          US-09-252-691-7056
Perfect score:  1165
Sequence:       1 AIMRQLITPENTMKTISFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters:   188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :  A_Geneseq_36:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	238	20.4	239	1	W80690	S. pneumoniae prot
2	128.5	11.0	148	1	W28082	Staphylococcus aur
3	88.5	7.6	302	1	W97722	Staphylococcus aur
4	85.5	7.3	591	1	R52825	DRAP-deaminase, DN
5	84	7.2	660	1	W93393	Human HEV ORF 2 pr
6	83	7.1	163	1	W80707	S. pneumoniae prot
7	83	7.1	660	1	W93391	Human HEV ORF 2 pr
8	83	7.1	660	1	W93394	Human HEV ORF 2 pr
9	81	7.0	392	1	R43074	Aspergillus niger
10	81	7.0	660	1	R14619	protein encoded by
11	80	6.9	2032	1	Y00238	Enterococcus faeca
12	80	6.9	2032	1	Y00240	Enterococcus faeca
13	80	6.9	2032	1	Y00242	Enterococcus faeca
14	79	6.8	245	1	W01626	Human fibroblast g
15	79	6.8	245	1	W29289	Human fibroblast g
16	79	6.8	245	1	W53026	Mouse fibroblast g
17	79	6.8	245	1	W53011	Fibroblast growth
18	79	6.8	525	1	R96103	Hepatitis E virus
19	79	6.8	540	1	R96101	Hepatitis E virus
20	79	6.8	547	1	R26189	Epidemic NANBH vir
21	79	6.8	549	1	R96091	Hepatitis E virus
22	79	6.8	549	1	W76367	Hepatitis E virus
23	79	6.8	660	1	R38785	HEV ORF2 protein.
24	79	6.8	660	1	R39306	Burma strain HEV O
25	79	6.8	660	1	R51265	HEV strain protein
26	79	6.8	660	1	R70023	Hepatitis E virus
27	79	6.8	660	1	R96089	Hepatitis E virus
28	79	6.8	660	1	R91814	Hepatitis E virus
29	79	6.8	660	1	W35826	Hepatitis E virus
30	79	6.8	660	1	W71210	Protein encoded by
31	79	6.8	660	1	W76369	Hepatitis E virus
32	79	6.8	660	1	W80197	Protein encoded by
33	79	6.8	660	1	W81520	Hepatitis E virus
34	79	6.8	660	1	W93386	Human HEV ORF 2 pr
35	79	6.8	660	1	W93388	Human HEV ORF 2 pr
36	79	6.8	660	1	W93389	Human HEV ORF 2 pr
37	79	6.8	660	1	W93392	Human HEV ORF 2 pr
38	79	6.8	660	1	W93395	Human HEV ORF 2 pr
39	78	6.7	5069	1	W52846	A. mediterranei ri
40	77.5	6.7	621	1	Y00243	Enterococcus faeca
41	77.5	6.7	2100	1	W89579	Calcium permeable
42	77	6.6	660	1	W93390	Human HEV ORF 2 pr
43	77	6.6	923	1	W96246	Human semaphorin r
44	76.5	6.6	1596	1	W31347	Rat tumour suppres
45	75.5	6.5	3418	1	W23287	Human breast and o

ALIGNMENTS

```
RESULT   1
W80690
ID  W80690 standard; Protein; 239 AA.
AC  W80690;
DT  24-DEC-1998  (first entry)
DE  S. pneumoniae protein of unknown function.
KW  Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
KW  virulence; antibody; infection; detection; treatment; hypothetical;
KW  cell wall biosynthetic; external target; minimal gene set protein.
OS  Streptococcus pneumoniae.
PN  WO9826072-A1.
PD  18-JUN-1998.
PF  09-DEC-1997; U22578.
PR  13-DEC-1996; US-036281.
PA  (ELIL ) LILLY & CO ELI.
PI  Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI  Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR,
PI  Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI  Young Bellido ML;
DR  WPI; 98-348529/30.
DR  N-PSDB; V65255.
PT  Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT  for evaluating gene expression, and identification of virulence
PT  genes
PS  Claim 3; Pages 278-279; 333pp; English.
CC  This sequence represents a Streptococcus pneumoniae protein of unknown
CC  function. The invention provides DNA sequences (V65201 to V65304) from
CC  the Streptococcus pneumoniae genome and corresponding protein sequences
CC  (W80605 to W80728). The protein sequences are classified as hypothetical,
CC  cell wall biosynthetic, external target, or minimal gene set proteins. A
CC  recombinant host containing a vector comprising any of the above nucleic
CC  acids can be used for the recombinant expression of the proteins. The
CC  invention also provides a DNA chip having arrayed on it at least 15 base
CC  pair fragment of any one or more of these DNA sequences. The DNA chip can
CC  be used methods for evaluating gene expression in S. pneumoniae and for
CC  identifying virulence genes in S. pneumoniae. Antibodies that selectively
CC  bind to the above proteins or peptide fragments can be used to treat
CC  S. pneumoniae infection. The antibodies can also be used to detect
CC  S. Pneumoniae cells.
SQ  Sequence 239 AA;

Query Match         20.4%;      Score 238;  DB 1;  Length 239;
Best Local Similarity    33.5%;      Pred. No. 2.1e-18;
Matches   60;  Conservative  32;  Mismatches  71;  Indels  16;  Gaps  4;

Qy   45  LFNKPYDVLPQFTDEAGRSTLKDFIP--VQGVYAAGRLRDSEGLLVLTNDGVLQARLTQ 102
             |||| |: |:  :|: :| |: |:  |||  :|: :| |:|| ||||||||
Db   63  LLNKPRGVISSVTDDKGRKTVVDLLPNVKERIYPVGRLDWDTSGVLILTNDGDFTDEMIH 122

Qy  103  PGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRER 162
         | :: | :|::       :|  ||: .|::|:|:| ||        :  |:::
Db  123  PRNEIDKVTVARVKGVANKDNLRPLTRGLEIDGKKTKPAVYEIL------KVDPVKNR 174

Qy  163  KSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRIRYAMGSYTLDSLANGEWRDVTPKE 221
             | :.::||| || |||::||: ||  | |  :|  .: ||     |||||||||||
Db  175  ----SVVQLTIHEGRNHQVKKMFEAVGLQVDKLSRTRFGHLDL-TLRPGESRRLNKKE 227
```

RESULT 2
W28082
```
ID   W28082 standard; Protein; 148 AA.
AC   W28082;
DT   01-SEP-1998 (first entry)
DE   Staphylococcus aureus protein of unknown function.
KW   Staphylococcus aureus protein; antisense sequence; control;
KW   Staphylococcus aureus protein; ribozyme; bacterial gene expression;
KW   vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome.
OS   Staphylococcus aureus.
PN   WO9730070-A1.
PD   21-AUG-1997.
PF   19-FEB-1997; U02318.
PR   20-FEB-1996; US-011888.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR   WPI; 97-424969/39.
DR   N-PSDB; T84021.
PT   Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT   to isolate antimicrobial compounds, and in vaccines against S.
PT   aureus infection
PS   Claim 6; Page 453; 989pp; English.
CC   The present sequence represents a Staphylococcus aureus protein of
CC   unknown function. The DNA sequence was isolated from a library of
CC   clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
CC   be used in the construction of ribozymes and antisense sequences to
CC   control the expression of Staphylococcal genes. The DNA sequence is
CC   also useful as a source of regulatory elements for the control of
CC   bacterial gene expression. The present protein may be used to produce
CC   vaccines to enable a host to produce specific antibodies with
CC   antibacterial action. These vaccines and antibodies would protect
CC   a host against invasion by S. aureus, and conditions relating to
CC   Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC   skin syndrome, and toxic shock syndrome.
SQ   Sequence 148 AA;

Query Match             11.0%;  Score 128.5;  DB 1;  Length 148;
Best Local Similarity   32.5%;  Pred. No. 1.2e-06;
Matches   40;  Conservative   15;  Mismatches   45;  Indels   23;  Gaps   6;

Qy   7 ITPENTMTKTSFRKHRVERFSSRQATRRFPEPQPTRVILFNKPYDVLPQFTDEAGRSTLK 66
         | ||| |:| :|    :|              |: ||| | |             |:
Db  41 IEPEN--DKITVRGELIEYIEN-----------------VYIMLNKPKGYI-SATEDRHSKTVI 84

Qy  67 DFIPVQ---GVYAAGRLRDRDSEGLLVLNTNDGVLQARLTQPGKRTGKIYYVQVEGEP---D 120
         | || :   :|||:|:||||||:||||  ||||||||| | || ||:|::| :| |
Db  85 DLIPEYQHLNIFPVGRLRDKDTEGLLLITNDGDFNHELMSPNKHVSKKIYEV-ISGNPITED 143

Qy 121 DAS 123
          |
Db 144 DNS 146
```

RESULT 3
W97722
```
ID   W97722 standard; Protein; 302 AA.
AC   W97722;
DT   21-MAY-1999 (first entry)
DE   Staphylococcus aureus mutant P15B9 virulence gene product.
KW   Virulence; P15B9; vaccine; antibacterial; antibiotic; screening;
KW   bacteraemia.
OS   Staphylococcus aureus.
PN   WO9901473-A2.
PD   14-JAN-1999.
PF   03-JUL-1998; G01974.
PR   03-JUL-1997; US-887534.
PA   (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
PI   Holden DW;
DR   WPI; 99-105999/09.
DR   N-PSDB; X07134.
PT   Inhibition of virulence genes from Staphylococcus aureus - useful
PT   for, e.g. screening for potential anti-microbial agents
PS   Claim 1; Page 179-180; 203pp; English.
CC   This is the deduced amino acid sequence of the protein product of a
CC   virulence gene (see X07134) newly identified in Staphylococcus aureus
CC   mutant P15B9. S. aureus mutants containing chromosomal insertions
CC   of a signature-tagged transposon were generated, and those with
CC   attenuated virulence were identified in a mouse model of bacteraemia.
CC   The nucleotide sequences of regions flanking the transposon insertion
CC   sites of these mutants were determined, and database sequence
CC   comparisons were performed to identify these virulence genes (see
CC   X07088-136) and the possible function of their protein products
CC   (see W97680-724). The P15B9 virulence gene product is unknown but
CC   shows 41% identity to the hypothetical 33.7 kDa protein YHCT of
CC   Bacillus subtilis p54604. A claimed method of identifying an
CC   antibacterial agent involves assaying potential agents for the
CC   ability to interfere with the expression of S. aureus virulence
CC   gene products. Also new is a S. aureus organism containing a
CC   functional mutation in one of the virulence genes, and its use in
CC   vaccine compositions.
SQ   Sequence 302 AA;

Query Match             7.6%;   Score 88.5;   DB 1;  Length 302;
Best Local Similarity   23.6%;  Pred. No. 0.09;
Matches   49;  Conservative   22;  Mismatches   72;  Indels   65;  Gaps   8;

Qy  39 QPTRVILFNKPYDVLPQFTDEAGRSTLKDFI----PVQG---VYAAGRLDRDSEGLLVL 90
         :| | ||  |:   :      | :|  :|: | :  |: :|
Db  88 EDNHMLIINKPAGIATHPNEDGQTGTLANLIAYHYQINGETCKVRHVHRLDQDTSGAIVF 147

Qy  91 TNDGVLQARLTQP-GKRTGKIYYVQV-EGEPDDASLAKLRNGVTLNDGPTLPAGIERVNE 148
         :|  |:|:  | |  ||: ||  |: | |:|
Db 148 AKHRLAHAILDQQLEKKTLKRYYTAIAEG-------KLR------------TKKGTI---- 186

Qy 149 PEWLWPRNPPIRERKSIPT------------------------SWLKITLYEGRNRQV 182
            |||| :||:   :| ::          || ||
Db 186 ------NPPIGRDRSHPTRRRVSPGGQTAVTHFKVMASNAKERLSLVELELETGRTHQI 238

Qy 183 RRMTAHVGFPTLRLIRYAMGSYTLDSLA 210
         ||| | :|  ||  | || | |  :||
Db 239 RVHLASLGHPLTGDSLIYGGGSKLLNRQA 266
```

RESULT 4
R52825
```
ID   R52825 standard; Protein; 591 AA.
AC   R52825;
DT   24-JAN-1995 (first entry)
DE   DRAP-deaminase.
KW   Riboflavin; expression; rib-1; guanosine triphosphate;
KW   GTP-cyclohydrolase II; rib-7; HTP-reductase; rib-2; DRAP-deaminase;
KW   rib-4; DMRL-synthase; rib-5; riboflavin-synthase; rib-3;
KW   DBP-synthase.
OS   Saccharomyces cerevisiae.
PN   DE4238904-A.
PD   26-MAY-1994.
PF   19-NOV-1992; 238904.
PR   19-NOV-1992; DE-239904.
PA   (BADI ) BASF AG.
PI   Bultrago Sema MJ,  Garcia Ramirez JJ,  Gonzales-Hernandez GA;
PI   Revuelta Doval JL,  Santos Garcia MA;
DR   WPI; 94-177113/22.
DR   N-PSDB; Q62811.
PT   DNA sequences and expression vectors encoding riboflavin genes -
PT   are useful for the prodn. of riboflavin in eukaryotes
PS   Claim 2; Page 8-11; 28pp; German.
CC   The riboflavin genes isolated from S. cerevisiae allow the
CC   prodn. of riboflavin in eukaryotes. The gene products are:
CC   - rib-1 gene: guanosine triphosphate (GTP)-cyclohydrolase II
CC   (transforming GTP to 2,5-diamino-6-ribosylamino-4-
```

CC   (3H)-pyrimidine-5-phosphate);
CC   - rib-7 gene: HTP-reductase (transforming 2,5-diamino-6-ribosylamino-4-
CC   (3H)-pyrimidine-5-phosphate into 2,5-diamino-ribitylamino-
CC   2,4-(1H,3H) pyrimidine-5-phosphate (DRAP));
CC   - rib-2 gene: (DRAP)-deaminase (formation of 5-amino-6-ribityl
CC   amino-2,4-(1H,3H)-pyrimidinedion);
CC   - rib-4 gene: 6,7-dimethyl-8-ribityllumazine (DMRL)-synthase;
CC   - rib-5 gene: riboflavin-synthase;
CC   - rib-3 gene: L-3,4-dihydroxy-2-butanone-4-phosphate (DBP)-synthase.
SQ   Sequence 591 AA;

Query Match   7.3%;   Score 85.5; DB 1; Length 591;
Best Local Similarity   21.8%;   Pred. No. 0.52;
Matches   47; Conservative   27; Mismatches   59; Indels   83; Gaps   11;

Qy   21 HRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDE--------------AGR--- 63
                |:| : ||||| ::|||:: :||
Db   147 HKVHRH-------EPPVT-----SKPIDIV--FEDEDILVIDKPSSIPVHPTGRYRF 189

Qy   63 STLKDFIPVQ--GVYAAGRLDRDSEGLLVLTNDGVLQARL-TQPGKRTGKIYYVQVEG 117
                |||||||| | | ::|||||: |||||| : | | :| |||:||:||||||||
Db   190 NTITKMLERQLGYSVHPCNRLDKPTSGLMFLAKTPLGADRMGDQMKAREVTKEYVARVKG 249

Qy   118 EPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPI---------------RERKS 164
                | || :| | | ||||||| ||||:||||||||||||::: ||||
Db   250 E--------------FPIGIVEVDKP--VRSVNPKVALNAVCEMSDENAKHAKT 287

Qy   165 I------------PTSWLKITLYEGRNRQVRRMTAHVGFP 192
                : | || |: |||||
Db   288 VFQRVSVDGQTSIVKCKPLTGRTHQIRVHLQYLGFP 323

RESULT 5
W93393
ID   W93393 standard; Protein; 660 AA.
AC   W93393;
DT   11-JUN-1999 (first entry)
DE   Human HEV ORF 2 protein from strain Hyderabad.
KW   Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KW   vaccine; immunise; infection; detection; diagnosis; prevention.
OS   Hepatitis E virus.
PN   WO9904029-A2.
PD   28-JAN-1999.
PF   17-JUL-1998; U14665.
PR   18-JUL-1997; US-053069.
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI   Emerson SU, Meng X, Purcell RH;
DR   WPI: 99-132270/11.
PT   New isolated swine hepatitis E virus - used to develop products for
PT   the diagnosis, prevention and treatment of hepatitis E virus
PT   infection in mammals, particularly humans
PS   Example 1; Fig 3A; 70pp; English.
CC   This invention describes a swine hepatitis E virus (HEV) and its natural
CC   mutants which are capable of cross-reacting with antibodies reactive
CC   with a human HEV strain or natural mutants. The HEV and the proteins
CC   can be used in vaccines for immunising against HEV infection. The swine
CC   HEV can also be used in humans to prevent possible infection by human HEV. The
CC   swine HEV can also be used as a therapeutic treatment for infection by
CC   other strains of HEV. The swine HEV can also be used for the production
CC   of antibodies which can be used in therapy, detection and diagnosis. The
CC   products can also be used for determining the susceptibility of cells or
CC   organs to infection with swine HEV. The swine HEV is particularly useful
CC   for the development of agents for the prevention, treatment and detection
CC   of human HEV because it is not a human virus and thus can be handled both
CC   experimentally and clinically without fear of severe infection and/or
CC   contamination.
SQ   Sequence 660 AA;

Query Match   7.2%;   Score 84; DB 1; Length 660;
Best Local Similarity   24.6%;   Pred. No. 0.89;
Matches   59; Conservative   24; Mismatches   24; Indels   99; Gaps   12;

Qy   7 ITPENTMKTSFRKHRVERFSS--RQATRR----TPEPQPTRVILFNKP-YDVLPQFTDE 59
               :||:|| ||:|| ::||
Db   311 LTPGNTNT------------RVSRYSSTARHRLRRGADGTAELTTAARFMKDLYFTSNGVGE 363

Qy   60 AGR---------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTOPGKRTGK 109
                ||| ||||||||||||
Db   364 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV-----SANGEPTVK 411

Qy   110 IYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPIRERKSIPTS- 169
               : ||: || |: || | || |:
Db   412 L-YTSVENAQQDKGIA-IPNDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 469

Qy   169 --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG------EWRDVT 218
                :| | :::|| || ||: :| || :||
Db   470 VLWLSLTAAEYDQSTYGSSTGPV----------YVSDSVTLVNVATGAQAVARSLDWTKVT 520

RESULT 6
W80707
ID   W80707 standard; Protein; 163 AA.
AC   W80707;
DT   24-DEC-1998 (first entry)
DE   S. pneumoniae protein of unknown function.
KW   Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
KW   virulence; antibody; infection; detection; treatment; hypothetical;
KW   cell wall biosynthetic; external target; minimal gene set protein.
OS   Streptococcus pneumoniae.
PN   WO986072-A1.
PD   18-JUN-1998.
PF   09-DEC-1997; U22578.
PR   13-DEC-1996; US-036281.
PA   (ELIL) LILLY & CO ELI.
PI   Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI   Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR,
PI   Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI   Young Bellido ML;
DR   WPI: 98-348529/30.
DR   N-PSDB: V65267.
PT   Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT   for evaluating gene expression, and identification of virulence
PT   genes
PS   Claim 3; Page 299; 333pp; English.
CC   This sequence represents a Streptococcus pneumoniae protein of unknown
CC   function. The invention provides DNA sequences (V65201 to.V65304) from
CC   the Streptococcus pneumoniae genome and corresponding protein sequences
CC   (W80005 to W80028). The protein sequences are classified as hypothetical,
CC   cell wall biosynthetic, external target, or minimal gene set proteins. A
CC   recombinant host containing a vector comprising any of the above nucleic
CC   acids can be used for the recombinant expression of the proteins. The
CC   invention also provides a DNA chip having arrayed on it at least 15 base
CC   pair fragment of any one or more of these DNA sequences. The DNA chip can
CC   be used methods for evaluating gene expression in S. pneumoniae and for
CC   identifying virulence genes in S. pneumoniae. Antibodies that selectively
CC   bind to the above proteins or peptide fragments can be used to treat
CC   S. pneumoniae infection. The antibodies can also be used to detect
CC   S. pneumoniae cells.
SQ   Sequence 163 AA;

Query Match   7.1%;   Score 83; DB 1; Length 163;
Best Local Similarity   23.4%;   Pred. No. 0.15;
Matches   33; Conservative   27; Mismatches   56; Indels   24; Gaps   6;

Qy   65 LKDFIPVQGVYAAG--RLDRDSEGLLVLTNDGVLQARLTQ-PGKRTGKIYYVQVEGE- 119
               :|| | || :||| ||||||||:: : : : | | ||
Db   24 IKDLSGINGVLRPGIVHRIDKDTSGLLMIAKNDDAHLVLAQELKDKKSLRKYWAIVHGNL 83

Qy   119 PDDASLAKLRNGVTLNDGPTLPAG-IERVNEPEWLWPRNPIRERKSI-----PTSWLKI 172
               :|:| || :|:
Db   84 PNDRGVIE-----------APIGRSEKDRKKQAVTAKGKPAVTRFHVLERFGDYSLVEL 131

Qy   173 TLYEGRNRQVRRMTAHVGFP 192

Db   132 QLETGCRTHQIRVHMAYIGHP 151
            | || || | |::| |

RESULT 7
ID W93391 standard; Protein; 660 AA.
AC W93391;
DT 11-JUN-1999 (first entry)
DE Human HEV ORF 2 protein from strain Uigh179.
KW Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KW vaccine; immunise; infection; detection; diagnosis; prevention.
OS Hepatitis E virus.
PN WO9904029-A2.
PD 28-JAN-1999.
PF 17-JUL-1998; U14665.
PR 18-JUL-1997; US-053069.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Meng X, Purcell RH;
DR WPI; 99-132270/11.
PT New isolated swine hepatitis E virus - used to develop products for
PT the diagnosis, prevention and treatment of hepatitis E virus
PT infection in mammals, particularly humans
PS Example 1; Fig 3A; 70pp; English.
CC This invention describes a swine hepatitis E virus (HEV) and its natural
CC mutants which are capable of cross-reacting with antibodies reactive
CC with a human HEV strain or natural mutants. The swine HEV and the proteins
CC can be used in vaccines for immunising against HEV infection. The swine
CC HEV can be used in humans to prevent possible infection by human HEV. The
CC swine HEV can also be used as a therapeutic treatment for infection by
CC other strains of HEV. The swine HEV can also be used for the production
CC of antibodies which can be used in therapy, detection and diagnosis. The
CC products can also be used for determining the susceptibility of cells or
CC organs to infection with swine HEV. The swine HEV is particularly useful
CC for the development of agents for the prevention, treatment and detection
CC of human HEV because it is not a human virus and thus can be handled both
CC experimentally and clinically without fear of severe infection and/or
CC contamination.
SQ Sequence 660 AA;

Query Match   7.1%; Score 83; DB 1; Length 660;
Best Local Similarity 24.6%; Pred. No. 1.2;
Matches 59; Conservative 25; Mismatches 98; Indels 58; Gaps 12;

Qy   7 ITPENTMTKTSFRKHRVERFSS--RQATRR----TPEPQPTRVLFNKP-YDVLPQFTDE 59
         :|| || || ||:|| | || ||| | || || :| ::|||
Db 311 LTPGNTNT-------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE 363

Qy   60 AGR---------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
         ||| || | :| | :||| | |:| |: | :| |: |||
Db 364 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SAHGEPTVK 411

Qy 110 IYYYQVGEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
         | | :| | :|| || | :|| |:| :| | ::||:
Db 412 L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 469

Qy 169 --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG-------EWRDVT 218
          | | || | | | | |::| |
Db 470 VLWLSLTAAEYDQSTYGSSTAPV---------YVSDSVTLVNVATGAQAVARSLDWTKVT 520

RESULT 9
ID R43074 standard; Protein; 392 AA.
AC R43074;
DT 26-APR-1994 (first entry)
DE Aspergillus niger Sulphydryl oxidase (SOX).
KW Genomic sequence; Sulphydryl oxidase; scavenger molecule;
KW disulphide bond formation; oxidation; free sulphydryl group.
OS Aspergillus niger.
FH Key   Location/Qualifiers
FT peptide   1..19
FT   /label= signal_peptide
FT protein   20..392
FT   /label= SOX
PN EP-565172-A.
PD 13-OCT-1993.
PF 31-MAR-1993; 200922.
PR 10-APR-1992; EP-201027.
PA (UNIL) QUEST INT BV.
PI Maat J, Musters W, Schaap PJ, Stam H, Van de Vondervoort PJI;
PI Verbakel JMA, Visser J;
DR WPI; 93-322408/41.
DR N-PSDB; Q50008.
PT DNA encoding sulphydryl oxidase - used for prodn. of enzyme,
PT useful in baking and for removing off-flavour from milk or beer
PS Claim 5; Fig 7; 42pp; English.

Db   132 QLETGCRTHQIRVHMAYIGHP 151
            | || || | |::| |

RESULT 8
ID W93394 standard; Protein; 660 AA.
AC W93394;
DT 11-JUN-1999 (first entry)
DE Human HEV ORF 2 protein from strain NEBL.
KW Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KW vaccine; immunise; infection; detection; diagnosis; prevention.
OS Hepatitis E virus.
PN WO9904029-A2.
PD 28-JAN-1999.
PF 17-JUL-1998; U14665.
PR 18-JUL-1997; US-053069.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Meng X, Purcell RH;
DR WPI; 99-132270/11.
PT New isolated swine hepatitis E virus - used to develop products for
PT the diagnosis, prevention and treatment of hepatitis E virus
PT infection in mammals, particularly humans
PS Example 1; Fig 3A; 70pp; English.
CC This invention describes a swine hepatitis E virus (HEV) and its natural
CC mutants which are capable of cross-reacting with antibodies reactive
CC with a human HEV strain or natural mutants. The swine HEV and the proteins
CC can be used in vaccines for immunising against HEV infection. The swine
CC HEV can be used in humans to prevent possible infection by human HEV. The
CC swine HEV can also be used as a therapeutic treatment for infection by
CC other strains of HEV. The swine HEV can also be used for the production
CC of antibodies which can be used in therapy, detection and diagnosis. The
CC products can also be used for determining the susceptibility of cells or
CC organs to infection with swine HEV. The swine HEV is particularly useful
CC for the development of agents for the prevention, treatment and detection
CC of human HEV because it is not a human virus and thus can be handled both
CC experimentally and clinically without fear of severe infection and/or
CC contamination.
SQ Sequence 660 AA;

Query Match   7.1%; Score 83; DB 1; Length 660;
Best Local Similarity 23.8%; Pred. No. 1.2;
Matches 57; Conservative 26; Mismatches 99; Indels 58; Gaps 11;

Qy   7 ITPENTMTKTSFRKHRVERFSS------RQATRRTPEPQPTRVLFNKP-YDVLPQFTDE 59
         :|| || || ||:|| | |: | | || || :| ::|||
Db 311 LTPGNTNT-------RVSRYSSTARHRLRRGTDGTAELTTAATRFMKDLYFTSTNGVGE 363

Qy   60 AGR---------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
         ||| || | :| | :||| | |:| |: | :| |: |||
Db 364 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SANGEPTVK 411

Qy 110 IYYYQVGEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
         | | :| | :|| || | :|| |:| :| | ::||:
Db 412 L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 469

Qy 169 --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG-------EWRDVT 218
          | | || | | | | |::| |
Db 470 VLWLSLTAAEYDQSTYGSSTGPV---------YVSDSVTLVNVATGAQAVARSLDWTKVT 520

```
CC   A PCR product amplified using primers based on partial peptide
CC   fragments of purified A.niger SOX was used to screen an A.niger
CC   N400 genomic DNA library in lambda-EMBL4. A 2.5kb EcoRI fragment
CC   comprising the entire SOX gene was identified in positive plaques
CC   and was sequenced. Additional upstream sequence information was
CC   obtained from a 6.4kb NsiI fragment also present in positive
CC   plaques. The SOX protein is encoded by 3 exons.
SQ   Sequence  392 AA;

Query Match               7.0%;   Score 81;   DB 1;   Length 392;
Best Local Similarity    25.5%;   Pred. No. 0.89;
Matches  42;   Conservative  28;   Mismatches  63;   Indels  32;   Gaps   9;

Qy     2  IMRQLITPENTMKTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAG   61
             ||||| ||| :: : ::||| :::::: :::
Db   258  IFRVHFTDGSSVVPNTF------ITNYPTAQRSTLPEELSLVMVDNKIDT---TDYTG   306

Qy    62  RSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDD 121
             | ||||         :|:|:      :      | :  :|:|  ||| :||
Db   307  MRT-----SLSGVYAVG--DCNSDG----STNVPHAMFS--GKRAGVYVHVEMSREESN 352

Qy   122  ASLAKLRNGVTLNDGPTLPAGIERV--NEPEWLWPRNPPIRERKS 164
             |:::|:|:  |   |  || |||||  | ||||| |:|
Db   353  AAISK-RD---FDRRALEKQTERMVGNEMEDLWKRVLENHHRRS 392


RESULT  10
R14619
ID   R14619 standard; Protein; 660 AA.
AC   R14619;
DT   16-JAN-1992  (first entry)
DE   Protein encoded by ORF 2 of Burmese ET-NANB viral strain.
KW   enterically transmitted non-A, non-B hepatitis virus; hepatitis C;
KW   HCV; E.coli strain BB4; ATCC deposit number 67717; Burma.
OS   Enterically transmitted non-A, non-B hepatitis virus.
PN   WO9115603-A.
PD   17-OCT-1991.
PF   05-APR-1991;  U02368.
PR   05-APR-1990;  US-505888.
PA   (GENE-) GENELABS INC.
PA   (USSH ) US DEPT HEALTH & HUMAN.
PI   Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;
PI   Fry KE;
DR   WPI; 91-325242/44.
DR   N-PSDB; Q14412.
PT   New viral proteins from non A-non-B hepatitis agent - used to
PT   treat and prevent enterically-transmitted non-A non-B hepatitis
PT   virus
PS   Disclosure; Page 19; 117pp; English.
CC   A positive clone ET1.1 was identified in a library prepared from
CC   bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.
CC   Both strands of ET11 were sequenced. Identity of the sequence with
CC   sequences in etiologic agents has been confirmed by locating a
CC   similar sequence in a viral strain isolated in Burma. This protein
CC   is encoded by the second longest ORF (ORF 2) of the Burma strain.
CC   (See Q14410 for ET1.1).
SQ   Sequence  660 AA;

Query Match               7.0%;   Score 81;   DB 1;   Length 660;
Best Local Similarity    24.2%;   Pred. No. 1.9;
Matches  58;   Conservative  25;   Mismatches  99;   Indels  58;   Gaps  12;

Qy     7  ITPENTMKTSFRKHRVERFSS--RQATRR----TPEPQPTRVILFNKP-YDVLPQFTDE   59
             :|| ||| ||| ||:::|| | ||| |:| |
Db   311  LTPGNTNT------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSNGVGE   363

Qy    60  AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK  109
             ||          |:|:|||  :||         |: |:| |  | :  :| | ||
Db   364  IGRGIALTLFNLADTLLGVLPTELISSAG------GQLFYSRPVV-----SANGEPTVK 411

Qy   110  IYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
```

```
Db   412  L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 469
             : ||   : || :  : :.  :: :: :: :
Qy   169  --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG------EWRDVT 218
             || ::|        :|:           ||  : : |:|
Db   470  VLWLSLTAAEYDQSTYGSSTGPV-------YVSDSVTLVNVATGAQAVARSLDWTKVT 520


RESULT  11
Y00238
ID   Y00238 standard; Protein; 2032 AA.
AC   Y00238;
DT   20-APR-1999  (first entry)
DE   Enterococcus faecalis protein EF123.
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
OS   Enterococcus faecalis.
PN   WO9850554-A2.
PD   12-NOV-1998.
PF   04-MAY-1998;  U08959.
PR   14-NOV-1997;  US-066009.
PR   06-MAY-1997;  US-044031.
PR   16-MAY-1997;  US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bailey C, Choi CH, Hromocky) A, Kunsch CA;
WPI; 99-070095/06.
DR   N-PSDB; X20228.
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
PS   Claim 9; Page 224-225; 301pp; English.
CC   The present sequence represents a protein isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
SQ   Sequence  2032 AA;

Query Match               6.9%;   Score 80;   DB 1;   Length 2032;
Best Local Similarity    24.5%;   Pred. No. 13;
Matches  48;   Conservative  19;   Mismatches  83;   Indels  46;   Gaps   9;

Qy    25  RFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRS---TLK-DFIPVQGVYAAGRL   80
             ||  |                           ::      :||  ||  ::
Db  1836  RFVETQA-------PTSYLLNETPSASFTIAKDNQGKPATVVLKAPFINYQGAAKLVKI 1887

Qy    81  DRDSEGL-----------LVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDD 121
             |:                LVLTNDGVLQARLRSDNQGLVQVNHLQPGKYT---FVETKA-PDG
Db  1888  DQQKNALAGAEFKVIDAETGGTVARSLRSDNQGLVQVNHLQPGKYT---FVETKA-PDG 1942

Qy   122  ASLAK----LRNGVTLNDGPTLPAGIERVNEPEWLWPR--NPPIRERKSIPTSWLKITL 174
             |:            :                       |:|   :|
Db  1943  YQLSKQAVAFTIAATAKDKPELVNAGTFVNEKQPVSKKTKPNQPTTKQAARETGWLGL-- 2001

Qy   175  YEGRNRQVRRMTAHVG 190
             ||
Db  2001  -PKTNTQVNFFVFIG 2015


RESULT  12
Y00240
ID   Y00240 standard; Protein; 2032 AA.
AC   Y00240;
DT   20-APR-1999  (first entry)
DE   Enterococcus faecalis protein EF124.
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
```

```
OS   Enterococcus faecalis.
PN   WO9850554-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; US08959.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bailey C, Choi GH, Hromockyj A, Kunsch CA;
DR   WPI; 99-070095/06.
DR   N-PSDB; X20230.
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
PS   Claim 9; Page 228-229; 301pp; English.
CC   The present sequence represents a protein isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
SQ   Sequence   2032 AA;

Query Match             6.9%;   Score 80;  DB 1;  Length 2032;
Best Local Similarity  24.5%;   Pred. No. 13;
Matches   48;  Conservative   19;  Mismatches   83;  Indels   46;  Gaps    9;

Qy    25  RFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRS---TLK-DFIPVQGVYAAGRL  80
              ||  ||  || |  | ||  || :  | :     ||   || || |
Db  1836  RFVETQA--------PTSYLLNETPSASFTIAKDNQGKPATVVLKAPFINYQGAAKLVKI 1887

Qy    81  DRDSEGL-----------LVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDD 121
          |:|           | | |:::|    |||||   :|: ||
Db  1888  DQQKNALAGAEFKVTDAETGQTVARSLRSDNQGLVQVNHLQPGKYT----FVETKA-PDG 1942

Qy   122  ASLAK----LRNGVTLNDGPTLPAGIERVNEPEWLWPR---NPPIRERKSIPTSWLKITL 174
             |:|    |     ||| |   ||| |   : |||    |||
Db  1943  YQLSKQAVAFTIAATAKDKPELVNAGTFVNEKQPVSKKTKPNQPTTKQAARETGWLGL-- 2001

Qy   175  YEGRNRQVRRMTAHVG 190
          |  || | |:
Db  2001  -PKTNTQVNYFVFIG 2015

RESULT 13
Y00242
ID   Y00242 standard; Protein; 2032 AA.
AC   Y00242;
DT   20-APR-1999 (first entry)
DE   Enterococcus faecalis protein EF125.
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
OS   Enterococcus faecalis.
PN   WO9850554-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; US08959.
PR   14-NOV-1997; US-066009.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bailey C, Choi GH, Hromockyj A, Kunsch CA;
DR   WPI; 99-070095/06.
DR   N-PSDB; X20232.
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
PS   Claim 9; Page 232; 301pp; English.
CC   The present sequence represents a protein isolated from
```

```
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
SQ   Sequence   2032 AA;

Query Match             6.9%;   Score 80;  DB 1;  Length 2032;
Best Local Similarity  24.5%;   Pred. No. 13;
Matches   48;  Conservative   19;  Mismatches   83;  Indels   46;  Gaps    9;

Qy    25  RFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRS---TLK-DFIPVQGVYAAGRL  80
              ||  ||  || |  | ||  || :  | :     ||   || || |
Db  1836  RFVETQA--------PTSYLLNETPSASFTIAKDNQGKPATVVLKAPFINYQGAAKLVKI 1887

Qy    81  DRDSEGL-----------LVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDD 121
          |:|           | | |:::|    |||||   :|: ||
Db  1888  DQQKNALAGAEFKVTDAETGQTVARSLRSDNQGLVQVNHLQPGKYT----FVETKA-PDG 1942

Qy   122  ASLAK----LRNGVTLNDGPTLPAGIERVNEPEWLWPR---NPPIRERKSIPTSWLKITL 174
             |:|    |     ||| |   ||| |   : |||    |||
Db  1943  YQLSKQAVAFTIAATAKDKPELVNAGTFVNEKQPVSKKTKPNQPTTKQAARETGWLGL-- 2001

Qy   175  YEGRNRQVRRMTAHVG 190
          |  || | |:
Db  2001  -PKTNTQVNYFVFIG 2015

RESULT 14
WO1626
ID   WO1626 standard; Protein; 245 AA.
AC   WO1626;
DT   21-APR-1997 (first entry)
DE   Human fibroblast growth factor homologous factor 2 (FHF-2).
KW   Fibroblast growth factor homologous factor 2; FHF-2;
KW   cell proliferation; differentiation; central nervous system;
KW   heart; neurodegenerative disease; neoplasm; cancer; cardiac disorder;
KW   stroke; Alzheimer's disease; Parkinson's disease; viral encephalitis;
KW   retinitis pigmentosa; macular degeneration; myocardial infarction;
KW   diagnosis; antibody; therapy.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   modified_site  242
FT                  /label= Glycosylation
FT                  /note= "potential N-glycosylation site"
PN   WO9635716-A1.
PD   14-NOV-1996.
PF   10-MAY-1996; US0665.
PR   12-MAY-1995; US-438439.
PA   (UXJO ) UNIV JOHNS HOPKINS SCHOOL MED.
PI   Macke JP, Nathans J,  Smallwood PM;
DR   WPI; 96-518616/51.
DR   N-PSDB; T58316.
PT   DNA encoding fibroblast growth factor homologous factor-2 - used to
PT   develop prods. for the diagnosis and treatment of cell proliferative
PT   disorders
PS   Claim 1; Page 26-28; 46pp; English.
CC   Human fibroblast growth factor homologous factor-2 (FHF-2) (WO1626)
CC   is a novel 30 kDa (SDS-PAGE) growth factor involved in the growth,
CC   survival and/or differentiation of cells within the central nervous
CC   system and in the heart. Its amino acid sequence was deduced from
CC   a clone (T58316) obtd. from a human retina cDNA library. FHF-2
CC   shows homology to other members of the FGF family. Recombinant
CC   FHF-2 can be produced in host cells and used in in vitro methods,
CC   e.g. for maintaining cells or tissues in culture prior to
CC   transplantation. Antibodies raised against FHF-2 can be used to
CC   treat cell proliferative disorders such as cancer, stroke,
CC   Alzheimer's disease, Parkinson's disease, retinitis pigmentosa,
```

Tue Jun 13 10:30:02 2000

us-09-252-691-7056.rag

Search completed: June 10, 2000, 09:37:10
Job time: 7288 sec

```
CC   viral encephalitis, macular degeneration or myocardial infarction.
SQ   Sequence   245 AA;

Query Match         6.8%;   Score 79;   DB 1;   Length 245;
Best Local Similarity  26.7%;   Pred. No. 0.74;
Matches  39;  Conservative  18;  Mismatches  45;  Indels  44;  Gaps  6;

Qy    15 KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF 56
         ||| |:   ||:| |::: || |||    :: :||
Db    35 KTSCDKNKLNVFSRVKLFGSKKRRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT 92

Qy    57 TDEAGRSTLKDFIPV------QGVYAAGRLDRDSEGLL------------VLTNDG 94
         || ||| ||  |||       |||  | |   :|||||   |       | |
Db    93 KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSELFTPECKFKESVFENYY 152

Qy    95 VLQARLTQPGKRTGKIYYYQV--EGE 118
         | |:  :   :::|| :||.: |||
Db   153 VTYSSMIYRQQQSGRGWYLGLNKEGE 178

RESULT  15
W29289
ID   W29289 standard; Protein; 245 AA.
AC   W29289;
DT   27-MAR-1998 (first entry)
DE   Human fibroblast growth factor homologous factor-2.
KW   Fibroblast growth factor homologous factor-2; FHF-2; FHF-3; human.
OS   Homo sapiens.
PN   WO9735007-A1.
PD   25-SEP-1997.
PF   21-MAR-1997; U04641.
PR   21-MAR-1996; US-621143.
PA   (UUJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI   Nathans J, Smallwood PM, Tong P;
DR   WPI; 97-480213/44.
DR   N-PSDB; T86315.
PT   Novel fibroblast growth factor homologous factor 3 polypeptide -
PT   useful for diagnosing and treating e.g. neurodegenerative diseases,
PT   trauma, stroke and cancer
PS   Example 1; Fig 7b; 79pp; English.
CC   This sequence comprises human fibroblast growth factor homologous
CC   factor-1 (FHF-2), a member of the fibroblast growth factor family.
CC   A cDNA clone (see T86315) for FHF-2 was newly isolated from a human
CC   retina cDNA on the basis of sequence similarity to FGF-1 (see
CC   W29288). FHF-1 and FHF-2 amino acid sequences were used to screen
CC   the GenBank database of conceptually translated DNA sequences
CC   (DBEST). A short region of human genomic DNA (DBEST accession
CC   number 76387) showed homology to approximately 25% of FHF-1 and
CC   FHF2. Primers based on sequence 76387 were used to amplify human
CC   genomic DNA and retina cDNA.  PCR products were then used as probes
CC   to isolate full-length cDNA clones (see T86313) for novel, claimed
CC   FHF-3 (see W29287), a cell growth, survival or differentiation
CC   factor.
SQ   Sequence   245 AA;

Query Match         6.8%;   Score 79;   DB 1;   Length 245;
Best Local Similarity  26.7%;   Pred. No. 0.74;
Matches  39;  Conservative  18;  Mismatches  45;  Indels  44;  Gaps  6;

Qy    15 KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF 56
         ||| |:   ||:| |::: || |||    :: :||
Db    35 KTSCDKNKLNVFSRVKLFGSKKRRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT 92

Qy    57 TDEAGRSTLKDFIPV------QGVYAAGRLDRDSEGLL------------VLTNDG 94
         || ||| ||  |||       |||  | |   :|||||   |       | |
Db    93 KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSELFTPECKFKESVFENYY 152

Qy    95 VLQARLTQPGKRTGKIYYYQV--EGE 118
         | |:  :   :::|| :||.: |||
Db   153 VTYSSMIYRQQQSGRGWYLGLNKEGE 178
```

us-09-252-691-7056.rag

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     June 10, 2000, 09:02:10 ; Search time 40.07 Seconds
                                       (without alignments)
                                       79.988 Million cell updates/sec

Title:           US-09-252-691-7056
Perfect score:   1165
Sequence:        1 AIMRQLITPENTMTKTSFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters:   145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :    Issued_Patents_AA:*
          1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/1/1aa/6_COMB.pep:*
          4: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
          5: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

          Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83.5	7.2	1820	3	US-07-998-289B-8	Sequence 8, Appli
2	82	7.0	245	3	US-08-705-245-2	Sequence 2, Appli
3	81	7.0	392	1	US-08-423-441-2	Sequence 2, Appli
4	79	6.8	245	2	US-08-438-439C-2	Sequence 24, Appl
5	79	6.8	245	3	US-08-951-822-24	Sequence 9, Appli
6	79	6.8	245	3	US-08-705-245-9	Sequence 11, Appl
7	79	6.8	245	4	PCT-US96-06665-2	Sequence 17, Appl
8	79	6.8	246	1	US-08-438-439C-11	Sequence 17, Appl
9	79	6.8	436	1	US-08-259-148A-17	Sequence 27, Appl
10	79	6.8	436	2	US-08-484-054-17	Sequence 25, Appl
11	79	6.8	436	2	US-07-876-941A-17	Sequence 15, Appl
12	79	6.8	525	4	PCT-US95-13703-27	Sequence 15, Appl
13	79	6.8	540	4	PCT-US95-13703-25	Sequence 19, Appl
14	79	6.8	549	4	PCT-US95-13703-15	Sequence 19, Appl
15	79	6.8	660	1	US-08-240-049B-15	Sequence 19, Appl
16	79	6.8	660	1	US-08-259-148A-19	Sequence 2, Appli
17	79	6.8	660	1	US-08-484-054-19	Sequence 2, Appli
18	79	6.8	660	2	US-07-876-941A-19	Sequence 13, Appl
19	79	6.8	660	2	US-07-876-941A-19	Sequence 23, Appl
20	79	6.8	660	4	PCT-US93-08849-2	Sequence 19, Appl
21	79	6.8	660	4	PCT-US95-13703-13	Sequence 2, Appli
22	77.5	6.7	2100	2	US-08-808-793-23	Sequence 7, Appli
23	77.5	6.7	2100	2	US-08-772-512A-19	Sequence 38, Appl
24	76.5	6.6	1346	2	US-08-635-121-2	Sequence 39, Appl
25	75.5	6.5	3418	3	US-08-603-753D-4	Sequence 41, Appl
26	74	6.4	263	3	US-08-483-857-7	
27	74	6.4	354	1	US-07-752-101A-38	
28	74	6.4	354	1	US-07-752-101A-39	
29	74	6.4	354	1	US-07-752-101A-41	
30	74	6.4	354	1	US-07-752-101A-51	Sequence 51, Appl
31	72	6.2	262	2	US-08-436-748-7	Sequence 7, Appl
32	72	6.1	413	2	US-08-960-756-2	Sequence 2, Appl
33	70.5	6.1	278	4	PCT-US94-03744-2	Sequence 2, Appl
34	70	6.0	525	4	PCT-US95-13703-28	Sequence 28, Appl
35	70	6.0	540	4	PCT-US95-13703-26	Sequence 26, Appl
36	70	6.0	549	4	PCT-US95-13703-16	Sequence 16, Appl
37	70	6.0	660	4	PCT-US95-13703-14	Sequence 14, Appl
38	69	5.9	609	1	US-08-716-301-4	Sequence 4, Appl
39	68.5	5.9	353	1	US-07-752-101A-35	Sequence 35, Appl
40	68.5	5.9	353	1	US-07-752-101A-36	Sequence 36, Appl
41	68.5	5.9	642	2	US-08-245-511-48	Sequence 48, Appl
42	68.5	5.9	642	2	US-08-600-993A-48	Sequence 48, Appl
43	68.5	5.9	693	2	US-08-380-403A-5	Sequence 2, Appl
44	68.5	5.9	693	2	US-08-380-403A-5	Sequence 5, Appl
45	68.5	5.9	693	2	US-08-895-628-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-07-998-289B-8
; Sequence 8, Application US/07998289B
; Patent No. 6027876
; GENERAL INFORMATION:
;   APPLICANT: Black, Bruce C
;   APPLICANT: Taylor, Martin
;   APPLICANT: Heckel, David G
;   TITLE OF INVENTION: Method for Monitoring Pesticide
;   TITLE OF INVENTION: Resistance
;   NUMBER OF SEQUENCES: 40
;   CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Darby & Darby PC
;    STREET: 805 Third Avenue
;    CITY: New York
;    STATE: New York
;    COUNTRY: US
;    ZIP: 10022
;   COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/07/998,289B
;    FILING DATE: 30-DEC-1992
;    CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;    NAME: Robinson, Joseph R
;    REGISTRATION NUMBER: 33,448
;    REFERENCE/DOCKET NUMBER: 0646/0A939
;   TELECOMMUNICATION INFORMATION:
;    TELEPHONE: 212-527-7700
;    TELEFAX: 212-753-6237
;    TELEX: 236687
;   INFORMATION FOR SEQ ID NO: 8:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 1820 amino acids
;    TYPE: amino acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-07-998-289B-8

Query Match             7.2%;  Score 83.5;  DB 3;  Length 1820;
Best Local Similarity  30.1%;  Pred. No. 1.5;
Matches  40;  Conservative  18;  Mismatches  50;  Indels  25;  Gaps  6;

Qy    11  NTMTKTSFRKHRVERFSSRQATRRTPEPQTRVILFNKPYDVLPQFTDEAGRSTLKD--F   68
          :||||:|.::|.:|| ||  |    |  | :: :||::|: :: ::|||:: ||||||  |
Db   646  STMTKESKLRNRNTRNQSVGATNGGTTCLDTNHKLDHRDYEIGLECTDEAGKIKHHDNPF  705
```

Qy   69 I-PVQGVYAAGRLDRDSEGLLVLTNDGVLQA--RLTQPGKKRTGKIYYVQVEGEPDDASLA 125
              | ||| | :: :|| || :|| : | :|| | |:
Db   706 IEPVQTQTVV----DMKDVMVL-NDIIEQAAGRHSRASDRGVSVYYFPPTEDDDE----- 755

Qy   126 KLRNGVTLNDGPT 138
              |||
Db   755 -------DGPT 758

RESULT 2
US-08-705-245-2
; Sequence 2, Application US/08705245
; Patent No. 6020189
; GENERAL INFORMATION:
;   APPLICANT: Nathans et al., Jeremy
;   TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR HOMOLOGOUS
;   TITLE OF INVENTION: FACTORS (FHFs) AND METHODS OF USE
;   NUMBER OF SEQUENCES: 37
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 4225 Executive Square, Suite 1400
;   CITY: La Jolla
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92037
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/705,245
;   FILING DATE: 30-AUG-1996
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Wetherell, Jr., John R.
;   REGISTRATION NUMBER: 31,678
;   REFERENCE/DOCKET NUMBER: 07265/094001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 619/678-5070
;   TELEFAX: 619/678-50999
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 245 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-705-245-2

Query Match   7.0%;   Score 82; DB 3; Length 245;
Best Local Similarity 26.7%;   Pred. No. 0.1;
Matches 39; Conservative 19; Mismatches 44; Indels 44; Gaps 6;

Qy   15 KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF 56
              ||| |: || :|| : || |||| |||
Db   35 KTSCDKNKLNVFSRVKLFGSKKRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT 92

Qy   57 TDEAGRSTLKDFIPV--------QGVYAAGRLDRDSEGLL-------------VLTNDG 94
              ||: | ||| ||| : ||| |
Db   93 KDEDSTYTLFNLIPVGLRVVAIGVQTKLYLAMNSEGYLYISELFTPECKFKESVFENYY 152

Qy   95 VLQARLTQPGKKRTGKIYYVQV--EGE 118
              |: : |||:: :|| :: : |
Db 153 VIYSSMIYRQQQSGRGWYLGLNKEGE 178

RESULT 3
US-08-423-441-2
; Sequence 2, Application US/08423441
; Patent No. 552926

; GENERAL INFORMATION:
;   APPLICANT: MAAT, JAN
;   APPLICANT: MUSTERS, WOUTER
;   APPLICANT: STAM, HEIN
;   APPLICANT: SCHAAP, PETER J.
;   APPLICANT: VAN DE VONDERVOORT, PETER J.
;   APPLICANT: VISSER, JACOB
;   APPLICANT: VERBAKEL, JOHANNES M.
;   TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA ENCODING A
;   TITLE OF INVENTION: RIPENING FORM OF A POLYPEPTIDE HAVING SULFHYDRYL OXIDASE
;   TITLE OF INVENTION: ACTIVITY
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
;   STREET: 1100 NEW YORK AVENUE, N.W.
;   CITY: WASHINGTON
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/423,441
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/044,620
;   FILING DATE: 09-APR-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: KOKULIS, PAUL N.
;   REGISTRATION NUMBER: 16773
;   REFERENCE/DOCKET NUMBER: 202744/T7019(V)
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-861-3000
;   TELEFAX: 202-822-0944
;   TELEX: 24453CUSH
;   TELEX: 24453CUSH
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 392 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-423-441-2

Query Match   7.0%;   Score 81; DB 1; Length 392;
Best Local Similarity 25.5%;   Pred. No. 0.28;
Matches 42; Conservative 28; Mismatches 63; Indels 32; Gaps 9;

Qy   2 IMRQLITPENTMTKTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAG 61
             | | || :::: || | | :: | : :|: | |: |
Db 258 IFRVHFTDGSSVVPNTF------ITNVPTAQRSTLPEELSLVMVDNKIDT----TDYTG 306

Qy   62 RSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDD 121
             || | ||: ||| : || | :: | ||:: :||
Db 307 MRT-----SLSGVYAVG--DCNSDG----STNVPHAMFS--GKRAGVYVHVEMSREESN 352

Qy 122 ASLAKLRNGVTLNDGPTLPAGIERV- NEPEWLWPRNPPIRERKS 164
             |::|| :| ||| || :
Db 353 AAISK-RD----FDRRALEKQTERMVGNEMEDLWKRVLENHHRRS 392

RESULT 4
US-08-438-439C-2
; Sequence 2, Application US/08438439C
; Patent No. 5876967
; GENERAL INFORMATION:
;   APPLICANT: Nathans, Jeremy
;   APPLICANT: Smallwood, Phillip M.
;   APPLICANT: Macke, Jennifer P.

```
; ORGANISM: Homo sapiens
US-08-951-822-24

Query Match                6.8%;    Score 79;  DB 2;  Length 245;
Best Local Similarity   26.7%;    Pred. No. 0.23;
Matches   39;   Conservative   18;   Mismatches   45;   Indels   44;   Gaps   6;

Qy     15  KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF  56
           |||  |:       ||:| : ||::: || ||||           | |
Db     35  KTSCDKNKLNVFSRVKLFGSKKRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT  92

Qy     57  TDEAGRSTLKDFIPV--------QGVYAAGRLDRDSEGLL----------VLTNDG  94
           ||  |||:|||      |||       |         |         |
Db     93  KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSELFTPECKFKESVFENYY 152

Qy     95  VLQARLTQPGKRTGKIYYVQV--EGE 118
           : |  : :|   ::|: :|: : |||
Db    153  VTYSSMIYRQQSGRGWYLGLNKEGE 178

RESULT  6
US-08-705-245-9
; Sequence 9, Application US/0870524S
; Patent No. 6020189
; GENERAL INFORMATION:
;   APPLICANT: Nathans et al., Jeremy
;   TITLE OF INVENTION:  FIBROBLAST GROWTH FACTOR HOMOLOGOUS
;   TITLE OF INVENTION:  FACTORS (FHFs) AND METHODS OF USE
;   NUMBER OF SEQUENCES:  37
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Fish & Richardson P.C.
;     STREET: 4225 Executive Square, Suite 1400
;     CITY: La Jolla
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 92037
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/705,245
;     FILING DATE: 30-AUG-1996
;     CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Wetherell, Jr., John R.
;     REGISTRATION NUMBER: 31,678
;     REFERENCE/DOCKET NUMBER: 07265/094001
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 619/678-5070
;     TELEFAX: 619/678-50999
; INFORMATION FOR SEQ ID NO:  9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 245 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:  not relevant
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-705-245-9

Query Match                6.8%;    Score 79;  DB 3;  Length 245;
Best Local Similarity   26.7%;    Pred. No. 0.23;
Matches   39;   Conservative   18;   Mismatches   45;   Indels   44;   Gaps   6;

Qy     15  KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF  56
           |||  |:       ||:| : ||::: || ||||           | |
Db     35  KTSCDKNKLNVFSRVKLFGSKKRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT  92

Qy     57  TDEAGRSTLKDFIPV--------QGVYAAGRLDRDSEGLL----------VLTNDG  94
           ||  |||:|||      |||       |         |         |
```
```
; TITLE OF INVENTION:  FIBROBLAST GROWTH FACTOR HOMOLOGOUS
; TITLE OF INVENTION:  FACTOR-2 AND METHODS OF USE
; NUMBER OF SEQUENCES:  25
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 4225 Executive Square, Suite 1400
;   CITY: La Jolla
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92037
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/438,439C
;   FILING DATE: May 12, 1995
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Haile, Lisa A.
;   REGISTRATION NUMBER: 38,347
;   REFERENCE/DOCKET NUMBER: 07265/046001
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 619/678-5070
;   TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 245 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-438-439C-2

Query Match                6.8%;    Score 79;  DB 2;  Length 245;
Best Local Similarity   26.7%;    Pred. No. 0.23;
Matches   39;   Conservative   18;   Mismatches   45;   Indels   44;   Gaps   6;

Qy     15  KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF  56
           |||  |:       ||:| : ||::: || ||||           | |
Db     35  KTSCDKNKLNVFSRVKLFGSKKRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT  92

Qy     57  TDEAGRSTLKDFIPV--------QGVYAAGRLDRDSEGLL----------VLTNDG  94
           ||  |||:|||      |||       |         |         |
Db     93  KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSELFTPECKFKESVFENYY 152

Qy     95  VLQARLTQPGKRTGKIYYVQV--EGE 118
           : |  : :|   ::|: :|: : |||
Db    153  VTYSSMIYRQQSGRGWYLGLNKEGE 178

RESULT  5
US-08-951-822-24
; Sequence 24, Application US/0895182214
; Patent No. 5989866
; GENERAL INFORMATION:
;   APPLICANT: Deisher, Theresa A.
;   APPLICANT: Conklin, Darrell C.
;   APPLICANT: Raymond, Fenella
;   APPLICANT: Bukowski, Thomas R.
;   APPLICANT: Holderman, Susan D.
;   APPLICANT: Hansen, Birgit
;   APPLICANT: Sheppard, Paul O.
;   TITLE OF INVENTION: NOVEL FGF HOMOLOGS
;   FILE REFERENCE: 96-20
;   CURRENT APPLICATION NUMBER: US/08/951,822A
;   CURRENT FILING DATE: 1997-10-16
;   NUMBER OF SEQ ID NOS: 36
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
;   LENGTH: 245
;   TYPE: PRT
```

```
Db    93 KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSEHFTPECKFKESVFENYY 152

Qy    95 VLQARLTQPGKRTGKIYYVQV--EGE 118
         |   | :::::| :: : |
Db   153 VTYSSMIYRQQSGRGWYLGLNKEGE 178
```

RESULT 7
PCT-US96-06665-2
; Sequence 2, Application PC/TUS9606665
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University School of Medicine
; TITLE OF INVENTION: FIBROBROBLAST GROWTH FACTOR HOMOLOGOUS
; TITLE OF INVENTION: FACTOR-2 AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06665
; FILING DATE: 10-MAY-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.,
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/046WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-06665-2

```
Query Match
Best Local Similarity    6.8%;    Score 79; DB 4; Length 245;
Matches    39; Conservative 18; Mismatches 45; Indels 44; Gaps 6;

Qy    15 KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF 56
         ||| ||:    ||:| | | || |||   | |
Db    35 KTSCDKNLNVFSRVKLFGSKKRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT 92

Qy    57 TDEAGRSTLKDFIPV-------QGVYAAGRLDRDSEGLL----------VLTNDG 94
         || |||||         :|||
Db    93 KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSELFTPECKFKESVFENY 152

Qy    95 VLQARLTQPGKRTGKIYYVQV--EGE 118
         |   | :::::| :: : |
Db   153 VTYSSMIYRQQSGRGWYLGLNKEGE 178
```

RESULT 8
US-08-438-439C-11
; Sequence 11, Application US/08438439C
; Patent No. 5876967
; GENERAL INFORMATION:
; APPLICANT: Nathans, Jeremy
; APPLICANT: Smallwood, Phillip M.
; APPLICANT: Macke, Jennifer P.
; TITLE OF INVENTION: FIBROBROBLAST GROWTH FACTOR HOMOLOGOUS

; TITLE OF INVENTION: FACTOR-2 AND METHODS OF USE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/438,439C
; FILING DATE: May 12, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/046001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-438-439C-11

```
Query Match
Best Local Similarity    6.8%;    Score 79; DB 2; Length 246;
Matches    39; Conservative 18; Mismatches 45; Indels 44; Gaps 6;

Qy    15 KTSFRKH------RVERFSSRQATRRTPEPQPTRVILFNKPYD----------VLPQF 56
         ||| ||:    ||:| | | || |||   | |
Db    35 KTSCDKNLNVFSRVKLFGSKKRRRRPEPQLKGIV--TKLYSRQGYHLQLQADGTIDGT 92

Qy    57 TDEAGRSTLKDFIPV-------QGVYAAGRLDRDSEGLL----------VLTNDG 94
         || |||||         :|||
Db    93 KDEDSTYTLFNLIPVGLRVVAIQGVQTKLYLAMNSEGYLYTSELFTPECKFKESVFENY 152

Qy    95 VLQARLTQPGKRTGKIYYVQV--EGE 118
         |   | :::::| :: : |
Db   153 VTYSSMIYRQQSGRGWYLGLNKEGE 178
```

RESULT 9
US-08-259-148A-17
; Sequence 17, Application US/08259148A
; Patent No. 5741490
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Bradley, Daniel W.
; APPLICANT: Twu, Jr-Shin
; APPLICANT: Purdy, Michael A.
; APPLICANT: Tam, Albert W.
; APPLICANT: Krawczynski, Krzysztof Z.
; APPLICANT: Yarbough, Patrice D.
; TITLE OF INVENTION: Hepatitis E Virus Vaccine and Method
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306

APPLICANT: Bradley, Daniel W.
APPLICANT: Twu, Jr-Shin
APPLICANT: Purdy, Michael A.
APPLICANT: Tam, Albert W.
APPLICANT: Krawczynski, Krzysztof Z.
TITLE OF INVENTION: Hepatitis E virus Vaccine and Method
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Dehlinger & Associates
   STREET: 350 Cambridge Avenue, Suite 250
   CITY: Palo Alto
   STATE: CA
   COUNTRY: USA
   ZIP: 94306
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/484,054
   FILING DATE: 07-JUN-1995
   CLASSIFICATION: 424
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 870,985
   FILING DATE: 20-APRIL-1992
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 822,335
   FILING DATE: 17-JAN-1992
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 505,888
   FILING DATE: 05-APRIL-1990
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 420,921
   FILING DATE: 13-OCTOBER-1989
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 367,486
   FILING DATE: 16-JUNE-1989
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 336,672
   FILING DATE: 11-APRIL-1989
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 208,997
   FILING DATE: 17-JUNE-1988
ATTORNEY/AGENT INFORMATION:
   NAME: Sholtz, Charles K.
   REGISTRATION NUMBER: 38,615
   REFERENCE/DOCKET NUMBER: 4600-0093.38
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (415) 324-0880
      TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 17:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 436 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
   MOLECULE TYPE: protein
   HYPOTHETICAL: NO
   ORIGINAL SOURCE:
      INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9
US-08-484-054-17

Query Match   6.8%; Score 79; DB 1; Length 436;
Best Local Similarity   24.2%; Pred. No. 0.56;
Matches   58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;

Qy   7 ITPENTMKTSFRKHRVERFSS--RQATRR----TPEPQPTRVILFNKP-YDVLPQFTDE 59
          :|| ||| | ||:|| || ||:|| || ||: | | :
Db   87 LTPGNTNT------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE 139

Qy   60 AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
          || || :| : : | | | | | :|| || |

COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/259,148A
   FILING DATE: 13-JUN-1994
   CLASSIFICATION: 424
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 822,335
   FILING DATE: 17-JAN-1992
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 420,921
   FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 367,486
   FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 336,672
   FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 208,997
   FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
   NAME: Sholtz, Charles K.
   REGISTRATION NUMBER: 38,615
   REFERENCE/DOCKET NUMBER: 4600-0093.20
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (415) 324-0880
      TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 17:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 436 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
   MOLECULE TYPE: protein
   HYPOTHETICAL: NO
   ORIGINAL SOURCE:
      INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9
US-08-259-148A-17

Query Match   6.8%; Score 79; DB 1; Length 436;
Best Local Similarity   24.2%; Pred. No. 0.56;
Matches   58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;

Qy   7 ITPENTMKTSFRKHRVERFSS--RQATRR----TPEPQPTRVILFNKP-YDVLPQFTDE 59
          :|| ||| | ||:|| || ||:|| || ||: | | :
Db   87 LTPGNTNT------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE 139

Qy   60 AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
          || || :| : : | | | | | :|| || |
Db 140 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SANGEPTVK 187

Qy 110 IYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
          : || : | :| | : :   :: :   : | :
Db 188 L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 245

Qy 169 --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG-------EWRDVT 218
          || :| :| | :: |:| :| | : | |
Db 246 VLWLSLTAAEYDQSTYGSSTGPV---------YVSDSVTLVNVATGAQAVARSLDWTKVT 296

RESULT 10
US-08-484-054-17
; Sequence 17, Application US/08484054
; Patent No. 5770689
; GENERAL INFORMATION:
   APPLICANT: Reyes, Gregory R.

US-07-876-941A-17

```
Query Match
Best Local Similarity   6.8%;   Score 79;   DB 2;  Length 436;
          Similarity 24.2%;   Pred. No. 0.56;
Matches  58;  Conservative  25;  Mismatches  99;  Indels  58;  Gaps  12;

Qy    7  ITPENTMTKTSFRKHRVERFSS--RQATRR---TPEPQPTRVILFNKP-YDVLPQFTDE  59
         ||  ||     ||  ||  ||  :||||        :|  ||| |  ||:| || ::|
Db   87  LITPGNNT-------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE  139

Qy   60  AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK  109
         ||          :|  : :||  |||  :  ||                |  |:  |
Db  140  IGRGIALTLFNLADTLLGGLPTELISSAG-------GQLFYSRPVV----SANGEPTVK  187

Qy  110  IYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS-  169
         :| |  ||: | |||  : | ||  ||  :   ||| |  |     ||  | : :
Db  188  L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND  245

Qy  169  --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG--------EWRDVT  218
           : :||           |  |||||||||||||||||||||||         ||||||
Db  246  VLWLSLTAAEYDQSTYGSSTGPV--------YVSDSVTLVNVATGAQAVARSLLDWTKVT  296
```

RESULT 11
US-07-876-941A-17
; Sequence 17, Application US/07876941A
; Patent No. 5885768
GENERAL INFORMATION:
~ APPLICANT: Reyes, Gregory R.
~ APPLICANT: Bradley, Daniel W.
~ APPLICANT: Tam, Albert W.
~ APPLICANT: Mitchell, Carl
~ TITLE OF INVENTION: Hepatitis E Virus Peptide Antigen and
~ TITLE OF INVENTION: Antibodies
~ NUMBER OF SEQUENCES: 76
~ CORRESPONDENCE ADDRESS:
~   ADDRESSEE: Dehlinger & Associates
~   STREET: 350 Cambridge Avenue, Suite 250
~   CITY: Palo Alto
~   STATE: CA
~   COUNTRY: USA
~   ZIP: 94306
~ COMPUTER READABLE FORM:
~   MEDIUM TYPE: Floppy disk
~   COMPUTER: IBM PC compatible
~   OPERATING SYSTEM: PC-DOS/MS-DOS
~   SOFTWARE: PatentIn Release #1.0, Version #1.25
~ CURRENT APPLICATION DATA:
~   APPLICATION NUMBER: US/07/876,941A
~   FILING DATE: 01-MAY-1992
~   CLASSIFICATION: 435
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 822,335
~   FILING DATE: 17-JAN-1992
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 505,888
~   FILING DATE: 05-APRIL-1990
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 420,921
~   FILING DATE: 13-OCTOBER-1989
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 367,486
~   FILING DATE: 16-JUNE-1989
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 336,672
~   FILING DATE: 11-APRIL-1989
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 208,997
~   FILING DATE: 17-JUNE-1988
~ ATTORNEY/AGENT INFORMATION:
~   NAME: Sholtz, Charles K.
~   REGISTRATION NUMBER: 38,615
~   REFERENCE/DOCKET NUMBER: 4600-0093.33
~ TELECOMMUNICATION INFORMATION:
~   TELEPHONE: (415) 324-0880
~   TELEFAX: (415) 324-0960
~ INFORMATION FOR SEQ ID NO: 17:
~   SEQUENCE CHARACTERISTICS:
~   LENGTH: 436 amino acids
~   TYPE: amino acid
~   TOPOLOGY: linear
~   MOLECULE TYPE: protein
~   HYPOTHETICAL: NO
~   ORIGINAL SOURCE:
~   INDIVIDUAL ISOLATE: C2, BURMA, FIGURE 9
```

RESULT 12
PCT-US95-13703-27
; Sequence 27, Application PC/TUS9513703
GENERAL INFORMATION:
~  APPLICANT:
~  TITLE OF INVENTION:  HEPATITIS E VIRUS ANTIGENS AND
~  TITLE OF INVENTION:  USES THEREFOR
~  NUMBER OF SEQUENCES:  31
~  CORRESPONDENCE ADDRESS:
~    ADDRESSEE:  Dehlinger & Associates
~    STREET:  P.O. Box 60850
~    CITY:  Palo Alto
~    STATE:  CA
~    COUNTRY:  USA
~    ZIP:  94306-0850
~  COMPUTER READABLE FORM:
~    MEDIUM TYPE:  Floppy disk
~    COMPUTER:  IBM PC compatible
~    OPERATING SYSTEM:  PC-DOS/MS-DOS
~    SOFTWARE:  PatentIn Release #1.0, Version #1.25
~  CURRENT APPLICATION DATA:
~    APPLICATION NUMBER:  PCT/US95/13703
~    FILING DATE:
~    CLASSIFICATION:
~  ATTORNEY/AGENT INFORMATION:
~    NAME:  Fabian, Gary R.
~    REGISTRATION NUMBER:  33,875
~    REFERENCE/DOCKET NUMBER:  4600-0293.41
~  TELECOMMUNICATION INFORMATION:
~    TELEPHONE:  (415) 324-0880
~    TELEFAX:  (415) 324-0960
~  INFORMATION FOR SEQ ID NO:  27:
~    SEQUENCE CHARACTERISTICS:
~      LENGTH:  525 amino acids
~      TYPE:  amino acid
~      STRANDEDNESS:  single
~      TOPOLOGY:  linear
~    MOLECULE TYPE:  protein
~    HYPOTHETICAL:  NO
~    ORIGINAL SOURCE:
~      INDIVIDUAL ISOLATE:  Hepatitis E virus (Burma strain)
~      INDIVIDUAL ISOLATE:  r62kDa, 56.5 kDa
PCT-US95-13703-27

```
Query Match
Best Local Similarity   6.8%;   Score 79;   DB 4;  Length 525;
          Similarity 24.2%;   Pred. No. 0.74;
Matches  58;  Conservative  25;  Mismatches  99;  Indels  58;  Gaps  12;
```

Qy   7 ITPENTMKTSFRKHRVERFSS--RQATRR---TPEPOPTRVILFNKP-YDVLPQFTDE 59
          ::|| || :|| | | || || :|:|| : |: || |:|| |
Db   200 LTPGNTNT------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE 252
Qy   60 AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
          || || | | :| | |: ||: |:||| | ||:
Db   253 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SANGEPTVK 300
Qy   110 IYYVQVEGEPPDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
          : | | : | : |||:|| || || ||||||:|||||||||||:|:||||||||
Db   301 L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 358
Qy   169 --WLKITLYEGRNRQVRRMTAHVGFPTLRIRYAMGSYTLDSLANG-------EWRDVT 218
            || :| || | | :   : | | ||||| ||
Db   359 VLWLSLTAAEYDQSTYGSSTGPV----------YSDSVTLVNVATGAQAVARSLDWTKVT 409

RESULT 13
PCT-US95-13703-25
; Sequence 25, Application PC/TUS9513703
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND
; TITLE OF INVENTION: USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13703
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0293.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hepatitis E virus (Burma strain)
; INDIVIDUAL ISOLATE: r62kDa, 58.1 kDa
PCT-US95-13703-25

Query Match   6.8%; Score 79; DB 4; Length 540;
Best Local Similarity 24.2%; Pred. No. 0.78;
Matches   58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;

Qy   7 ITPENTMKTSFRKHRVERFSS--RQATRR---TPEPOPTRVILFNKP-YDVLPQFTDE 59
          ::|| || :|| | | || || :|:|| : |: || |:|| |
Db   200 LTPGNTNT------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE 252
Qy   60 AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
          || || | | :| | |: ||: |:||| | ||:
Db   253 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SANGEPTVK 300

Qy   110 IYYVQVEGEPPDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
          : | | : | : |||:|| || || ||||||:|||||||||||:|:||||||||
Db   301 L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 358
Qy   169 --WLKITLYEGRNRQVRRMTAHVGFPTLRIRYAMGSYTLDSLANG-------EWRDVT 218
            || :| || | | :   : | | ||||| ||
Db   359 VLWLSLTAAEYDQSTYGSSTGPV----------YSDSVTLVNVATGAQAVARSLDWTKVT 409

RESULT 14
PCT-US95-13703-15
; Sequence 15, Application PC/TUS913703
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HEPATITIS E VIRUS ANTIGENS AND
; TITLE OF INVENTION: USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13703
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0293.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Hepatitis E virus (Burma strain)
; INDIVIDUAL ISOLATE: r62kDa, FIGURE 4
PCT-US95-13703-15

Query Match   6.8%; Score 79; DB 4; Length 549;
Best Local Similarity 24.2%; Pred. No. 0.8;
Matches   58; Conservative 25; Mismatches 99; Indels 58; Gaps 12;

Qy   7 ITPENTMKTISFRKHVERFSS--RQATRR---TPEPOPTRVILFNKP-YDVLPQFTDE 59
          ::|| || :|| | | || || :|:|| : |: || |:|| |
Db   200 LTPGNTNT------RVSRYSSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE 252
Qy   60 AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
          || || | | :| | |: ||: |:||| | ||:
Db   253 IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SANGEPTVK 300
Qy   110 IYYVQVEGEPPDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS- 169
          : | | : | : |||:|| || || ||||||:|||||||||||:|:||||||||
Db   301 L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND 358
Qy   169 --WLKITLYEGRNRQVRRMTAHVGFPTLRIRYAMGSYTLDSLANG-------EWRDVT 218
            || :| || | | :   : | | ||||| ||

Db   359   VLWLSLTAAEYDQSTYGSSTGPV--------YVSDSVTILVNVATGAQAVARSLDWTKVT   409

RESULT 15
US-08-240-049B-15
; Sequence 15, Application US/08240049B
; Patent No. 5686239
; GENERAL INFORMATION:
;   APPLICANT: Reyes, Gregory R.
;   APPLICANT: Tam, Albert W.
;   APPLICANT: Yarbough, Patrice O.
;   TITLE OF INVENTION: Hepatitis E Virus Peptides and Methods
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genelabs Technologies, Inc.
;   STREET: 505 Penobscot Drive
;   CITY: Redwood City
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94063
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/240,049B
;   FILING DATE: 09-MAY-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Charles K. Sholtz
;   REGISTRATION NUMBER: 38,615
;   REFERENCE/DOCKET NUMBER: 4600-0093.35; G32P4
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 324-0880
;   TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 15:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 660 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   INDIVIDUAL ISOLATE: Hepatitis E Virus (Burma strain)
;   INDIVIDUAL ISOLATE: ORF-2
US-08-240-049B-15

Query Match   6.8%;   Score 79;   DB 1;   Length 660;
Best Local Similarity   24.2%;   Pred. No. 1.1;
Matches   58;   Conservative   25;   Mismatches   99;   Indels   58;   Gaps   12;

Qy   7   ITPENTMTKTSFRKHRVERFSS--RQATRR----TPEPQPTRVILFNKP-YDVLFQFTDE   59
         :|| || :|| :|| : | |:||| | :|| |||
Db   311   LTPGNTNT------RVSRISSTARHRLRRGADGTAELTTAATRFMKDLYFTSTNGVGE   363

Qy   60   AGR--------STLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGK   109
         | : |: :|| : |: :   :|:|
Db   364   IGRGIALTLFNLADTLLGGLPTELISSAG------GQLFYSRPVV----SANGEPTVK   411

Qy   110   IYYVQVQEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTS-   169
         :| |: :| : | :| : || : :
Db   412   L-YTSVENAQQDKGIA-IPHDIDLGESRVVIQDYDNQHEQDRPTPSPAPSRPFSVLRAND   469

Qy   169   --WLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANG------.-EWRDVT   218
         !! :|| || ||| :|| :| || :| ||
Db   470   VLWLSLTAAEYDQSTYGSSTGPV--------YVSDSVTILVNVATGAQAVARSLDWTKVT   520

Search completed: June 10, 2000, 10:45:05
Job time: 6175 sec

OM protein - protein search, using sw model

```
                      GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:    June 10, 2000, 09:04:36 ; Search time 55.25 Seconds
                                       (without alignments)
                                       235.580 Million cell updates/sec

Title:     US-09-252-691-7056
Perfect score:   1165
Sequence:  1 AIMRQLITPENTMTKTSFRK........SYTLDSLANGEWRDVTPKEN 222

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters:        168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :       PIR_63.*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	942	80.9	207	2	D64858	probable pseudouri
2	598	51.3	240	2	I64156	probable pseudouri
3	389	33.4	185	2	S74428	probable pseudouri
4	298	25.6	249	2	S76104	hypothetical prote
5	264.0	22.7	239	2	G72400	16S pseudouridylat
6	262	22.5	257	2	D75462	ribosomal large su
7	259	22.2	249	2	B70427	conserved hypothet
8	253	21.7	229	2	S45555	conserved hypothet
9	245	21.0	291	2	H64874	probable pseudouri
10	244.5	21.0	357	2	A64169	hypothetical prote
11	233.5	20.0	254	2	F70504	hypothetical prote
12	215.5	18.5	371	2	T36871	probable pseudouri
13	214	18.4	186	2	S72955	hypothetical prote
14	209.5	18.0	235	2	G71658	hypothetical prote
15	206	17.7	231	2	F64987	hypothetical 25.9
16	201.5	17.3	290	2	E65209	hypothetical 32.5
17	200.5	17.2	238	2	A70350	conserved hypothet
18	198.5	17.0	157	2	H70004	conserved hypothet
19	186	16.0	232	2	F64169	hypothetical prote
20	178.5	15.3	235	2	H72025	probable pseudouri
21	176	15.1	338	2	C71603	pseudouridine synt
22	175.5	15.1	258	2	S36431	hypothetical prote
23	153.5	13.2	241	2	F71478	probable pseudouri
24	151	13.0	249	2	A70116	probable pseudouri
25	120	10.3	261	2	E71320	conserved hypothet
26	119.5	10.3	316	2	D70451	conserved hypothet
27	113.5	9.7	262	2	C64702	conserved hypothet
28	112.5	9.7	262	2	B71817	hypothetical prote
29	112.5	9.7	304	2	D72316	ribosomal large su
30	97	8.3	326	2	E65037	ftsH suppressor pr
31	96.5	8.3	327	2	S76181	hypothetical prote
32	95.5	8.2	217	2	C70533	probable lexA prot
33	90.5	7.8	309	2	S73869	hypothetical prote
34	88.5	7.6	302	2	H69823	conserved hypothet
35	88.5	7.6	454	2	T08280	hypothetical prote
36	88	7.6	303	2	A71557	probable pseudouri
37	88	7.6	366	2	S76662	hypothetical prote
38	86.5	7.4	1275	2	T18556	o-antigen biosynth
39	85.5	7.3	591	2	S50972	RIB2 protein - yea
40	85	7.3	376	2	S45107	hypothetical prote
41	85	7.3	404	2	S64480	hypothetical prote
42	84.5	7.3	946	2	S71168	Ca2+-transporting
43	84.5	7.3	1034	2	S60051	sodium channel alp
44	84.5	7.3	1034	2	S60060	sodium channel alp
45	84	7.2	324	2	F64144	hypothetical prote

ALIGNMENTS

```
RESULT 1
D64858
probable pseudouridylate synthase ymfC - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: D64858
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: D64858
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues:1-207 from <BLAT>
A;Cross-references: GB:AE000213; GB:U00096; NID:g1787371; PIDN:AAC74219.1; PID:g17873
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ymfC

Query Match             80.9%;  Score 942;  DB 2;  Length 207;
Best Local Similarity   86.4%;  Pred. No. 1e-76;
Matches  178;  Conservative  12;  Mismatches  16;  Indels  0;  Gaps  0;

Qy   13 MTKTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKDFIPVQ 72
          !  |||||  ||:|: |:!!:|::!| !|||||:||||||||||||||||||||||| |||
Db    1 MQKTSFRNHQVKRFSSQRSTRRKPENQPTRVILFNKPYDVLPQFTDEAGRKTLKEFIPVQ 60

Qy   73 GVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVT 132
          |||||||||||||||||||:|  : |||||||||||||||||| |  |||||
Db   61 GVYAAGRLDRDSEGLLVLTNNGALQARLTQPGKRTGKIYYVQVEGIPTQDALEALRNGVT 120

Qy  133 LNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFP 192
          ||:||||||| |:!||:!|||||||||||||||||||||||||||||||||||||
Db  121 LNDGPTLPAGAELVDEPAWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFP 180

Qy  193 TLRLIRYAMGSYTLDSLANGEWRDVT 218
          :|||||||||||||:||||||||:|:|
Db  181 TLRLIRYAMGDYSLDNLANGEWREVT 206

RESULT 2
I64156
probable pseudouridylate synthase HI0694 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C;Accession: I64156
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
Science 269, 495-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
```

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: I64156
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-240 <TIGR>
A;Cross-references: GB:U32752; GB:L42023; NID:g1573692; PIDN:AAC22354.1; PID:g1573697;
C;Superfamily: conserved hypothetical protein HI1243

Query Match   51.3%;   Score 598;   DB 2;   Length 240;
Best Local Similarity   59.2%;   Pred. No. 5.3e-46;
Matches 119; Conservative   29; Mismatches   49; Indels   4; Gaps   2;

Qy   15 KISFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKDFIPVQGV 74
           : | :::: |||: | :::| |||||| ||| |||| ||||||||||||:|| |
Db   28 KSAVNFHRTFKPKLKPKTLSLDE--TKVVLFNKPFDVLTQFTDEQGRATLKDFISIPNV 84

Qy   75 YAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIIYYVQVEGEPDDASLAKLRNGVTLN 134
           ||||||||||||||||| ||||:|||||||||||||||||||||||||||||||||||||
Db   85 YAAGRLDRDSEGLLILTNNGELQHRLADPKFKTEKTYWVQVEGIPEETDLAQLRKGVELK 144

Qy   135 DGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTL 194
           || | | : ::| ||||||||||||||||| ||||| || | |||||:||||||||||
Db   145 DGVTKSAKVRLISEPN-LWERNPPIRERKNIPTSWLEIKISEGRNRQVRRMTAHIGFPTL 203

Qy   195 RLIRYAMGSYTLDSLANGEWR 215
           ||| | :||::| :| :| :|
Db   204 RLVRVSMGLLSINGLENGSFR 224

RESULT 3
S74428
probable pseudouridylate synthase slr0612 - Synechocystis sp. (PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S74428
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74428
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-185 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAA16580.1; PID:d101731
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

Query Match   33.4%;   Score 389;   DB 2;   Length 185;
Best Local Similarity   52.8%;   Pred. No. 1.5e-27;
Matches   85; Conservative   16; Mismatches   42; Indels   18; Gaps   2;

Qy   78 GRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIIYYVQVEGEPDDASLAKLRNGVTLNDGP 137
           |||||||||||| | ::| :| : ||| || ||||||| :: | || ||| | | |
Db   2 GRLDQDSEGLLLLTSNGKLQHRLAHREFAHQRTYFAQVEGSPTDEDLEPLRRGITFADYP 61

Qy   138 TLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLI 197
           | | |||||||||||||||| ||||| |||| | |||||||||||||:| ||||||| :
Db   62 TRPAIAKIITEPDFP-PRNPPIRYRASIPTSWLSITLTEGRNRQVRRMTAAVGFPTLRLV 120

Qy   198 RY---------AMGSYTLDSLANGEWRDVTPKE 221
           || | :||:| :| || | : |:|
Db   121 RVQIQVTGRSPQQGKGKSAATWCLTLEGLSPGQWRPLIPWE 161

RESULT 4
S76104

hypothetical protein - Synechocystis sp.
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 29-Sep-1999
C;Accession: S76104
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-249 <KAN>
A;Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10082.1; PID:d101
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: conserved hypothetical protein HI1243

Query Match   25.6%;   Score 298;   DB 2;   Length 249;
Best Local Similarity   38.5%;   Pred. No. 2.8e-19;
Matches   72; Conservative   28; Mismatches   63; Indels   24; Gaps   3;

Qy   44 ILFNKPYDVLPQFTDEAGRSTLKDFIPV-----QGVYAAGRLDRDSEGLLVLTNDGVLQA 98
           : |||| |:| : :| |||||||:| | | ||||||||||:||:|||| |
Db   67 LLVNKPRDVLSTCDDPRGRKTVLDLLPQDLQRGKGLHPVGRLDRNSTGALLTNDGELTL 126

Qy   99 RLTQPGKRTGKIIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPP 158
           ||| |||||||||||||||||| ::|| || ||||| |||||||||||||:|:|::|| |
Db   127 RLTHPRYHLPKTYDVWLEGNPSDEDLEKWRSGMLDGKKTLPATLEVISENK--------- 179

Qy   159 IRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDS----LANGEW 214
           | :|| |||||:|::||:| :|| : | : :||| ||| ::: ::
Db   179 ------DQIHLLVTLTEGRNRQIRRLAEELGLTVLKLHRRTIGPLQLHTRGKVLGSGQF 231

Qy   215 RDVTPKE 221
           | :||:|
Db   232 RFLSPAE 238

RESULT 5
G72400
16S pseudouridylate synthase - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 26-Aug-1999
C;Accession: G72400
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316
A;Accession: G72400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <ARN>
A;Cross-references: GB:AE001708; GB:AE000512; NID:g4980740; PID:g4980760; TIGR:TM0264
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0264
C;Superfamily: conserved hypothetical protein HI1243

Query Match   22.7%;   Score 264.5;   DB 2;   Length 239;
Best Local Similarity   35.8%;   Pred. No. 2.6e-16;
Matches   64; Conservative   31; Mismatches   65; Indels   19; Gaps   3;

Qy   44 ILFNKPYDVLPQFTDEAGRSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQP 103
           : |||: | : :| |||||||||||||||: ||||||||||:|||||:||: | |
Db   62 ILFVKPSGVTVSTKDPHSETIMEFLPPLKGIFPVGRLDKDAEGLLIITNDGDFAHRVISP 121

Qy   104 GKRTGKIIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERK 163

```
Db  122 KWSVEREYIVKVEGEVTEDKIEKLKNGVTLRDG--FFAKAKRV-----------E 163
            ||  | |:|||   : ||:||||| ||         :  ||
Qy  164 SIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTL-DSLANGEWRDVTPKE 221
            |||  | :|:|| ::|||| | |  ::||:  ||||||:||||      ||:| :: :|
Db  164 KLSNDTLKIVITEGKYHQIKRMTAAVGLKTVHLKRTRIGGLVLPDDMKPGEYRFLSEEE 222
```

RESULT   6
D75462
ribosomal large subunit pseudouridine synthase B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: D75462
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250
A;Accession: D75462
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-257 <WHI>
A;Cross-references: GB:AE001942; GB:AE001942; GB:AE00513; NID:g6458611; PIDN:AAF10472.1; PID:g645861
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0896
A;Map position: 1

```
Query Match                    22.5%;   Score 262;   DB 2;   Length 257;
Best Local Similarity   35.8%;   Pred. No. 4.7e-16;
Matches   67;  Conservative  23;  Mismatches  73;  Indels  24;  Gaps   4;

Qy   38 PQPTRVILFNKPYDVLPQFTDEAGRSTLKDFIP-VQGVYAAGRLDRDSEGLLVLTNDGVL 96
            |  |||| | | : || | ||:: ||| ||||  ||| |||||:||||||||||||||||
Db   61 PETVTYALY-KPVGFVTTAHDEYGRRNVLDAMPDVPGLHPVGRLDKDSEGLLLLTNDGDL 119
Qy   97 QARLTQPGKRTGKIYYVQVEGE--PDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWP 154
            |||  ||| | ::||   ||:    :|||||:||||||| | ||::||:|:|||  :||
Db  120 TLTLHPRYGHEKAYRAWTEGREPPTQAELDVLVRGIAMDDGPAQALSAAPAEDGAY--- 177
Qy  155 RNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGEW 214
            |  ||||||| | ||| |  :| :: ||  : |  : || || |:|:|||  ::|  |
Db  177 -------------VVLGEGRNRQVRRMLEALGHPVGRLVRYRVGGLWLGDLNPGEY 219
Qy  215 RDVTPKE 221
            |:|:  ||
Db  220 RELGPRD 226
```

RESULT   7
B70427
conserved hypothetical protein aq_1464 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 29-Sep-1999
C;Accession: B70427
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: B70427
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-249 <AQF>
A;Cross-references: GB:AE000741; NID:g2983841; PIDN:AAC07410.1; PID:g2983855; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1464
C;Superfamily: conserved hypothetical protein HI1243

```
Query Match                    22.2%;   Score 259;   DB 2;   Length 249;
Best Local Similarity   34.6%;   Pred. No. 8.3e-16;
Matches   65;  Conservative  28;  Mismatches  77;  Indels  18;  Gaps   3;

Qy   37 EPQPTRVILFNKPYDVLPQF-TDEAGRSTLKDFIP--VQGVYAAGRLDRDSEGLLVLTND 93
            :  |||| | | | || |   |||:| | |||  |   |||| |||:||||||||||||:
Db   59 KPQRKRRIILNKPCCYLTQLGLQGRSPDGRKTIEELIKDIPERVFPVGRLDYNTEGLLILTND 118
Qy   94 GVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLW 153
            :|  |: || ||| | ::|:    |: ||:||:||:|:|| :||| |||||:||||||||
Db  119 GELANRILHPRYKLPKVYLALVEGKVDQKTLKRMKQGIELEDGFAKPDNIRIV------- 172
Qy  154 PRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGE 213
            | | ||||||: | ||| |  || :: || :  :|  | ||  || |:|:|||:| : : ::
Db  172 ----------RYEGKNTLLEITFHEGRKHLVKRFLGAFGHKVKRLKRIAIGPIKLGKLSPGK 223
Qy  214 WRDVTPKE 221
            ||:| |
Db  224 WRELNQGE 231
```

RESULT   8
S45555
conserved hypothetical protein ypuL - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 29-Sep-1999
C;Accession: S45555; A69943
R;Sorokin, A.; Zumstein, E.; Azevedo, V.; Ehrlich, S.D.; Serror, P.
submitted to the EMBL Data Library, November 1993
A;Reference number: S45533
A;Accession: S45555
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <SOR>
A;Cross-references: EMBL:L09228; NID:g410114; PIDN:g410137
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, H.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: A69943
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-229 <KUN>
A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14248.1; PID:e11855
A;Experimental source: strain 168
C;Genetics:
A;Gene: ypuL
C;Superfamily: conserved hypothetical protein HI1243

```
Query Match                    21.7%;   Score 253;   DB 2;   Length 229;
Best Local Similarity   35.2%;   Pred. No. 2.5e-15;
Matches   64;  Conservative  31;  Mismatches  69;  Indels  18;  Gaps   5;

Qy   39 QPTRVILFNKPYDVLPQFTDEAGRSTLKDF---IPVQGVYAAGRLDRDSEGLLVLTNDGV 95
            : : | | | :|| ||| |||||:|||:  :||||: |||| |||||||||||||||||||:|
Db   59 EPYYFLLY-KPRGVISAAQDDKGRKVVTDFFKNIP-QRIYPIGRLDYDTSGLLLLTNDGE 116
Qy   96 LQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPR 155
            : :||||||||| | ::| :   |  :| || |: || || :||||:||||| ||  ::
```

```
Db    117 FANKLMHPKYEIDKTYVAKVKGIPPKELLRKLERGIRLEEGKTAPAKAKLLS-------- 169
```

```
Query Match             21.0%;    Score 244.5;   DB 2;   Length 357;
Best Local Similarity   37.2%;    Pred. No. 2.6e-14;
Matches 67;  Conservative  24;  Mismatches  70;  Indels  19;  Gaps  5;
```

```
Qy     42 RVILFNKPYDVLPQFTDEAGRSTLKDFIP-VQGV--YAAGRLDRDSEGLLVLTNDGVLQA 98
Db    141 RVLMYYKPEGELCTRSDPEGRATVFDRLPRLTGSWIAVGRLDINTSGLLFTTDGELAN 200

Qy     99 RLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPP 158
Db    201 RLMHPSREVEREYSVRVFGQVDDAMLARLRKGVQLEDG-------LANFKEIKFTGVG 252

Qy    159 IRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTL-DSLANGEWRDV 217
Db    253 INQ-------WYDVTLMEGRNREVRRLWESQGIQVSRLIRIRYGNIKLMKGLPRGGWEEM 305
```

RESULT 11
F70504
hypothetical protein Rv1711 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 08-Oct-1999
C;Accession: F70504
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: F70504
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-254 <COL>
A;Cross-references: GB:Z98268; GB:AL123456; NID:g3261839; PIDN:CAB10968.1; PID:e33279
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1711

```
Query Match             20.0%;    Score 233.5;   DB 2;   Length 254;
Best Local Similarity   32.5%;    Pred. No. 1.6e-13;
Matches 68;  Conservative  25;  Mismatches  81;  Indels  35;  Gaps  5;
```

```
Qy     32 TRRTPEPQPTRV---------------ILFNKPYDVLPQFTDEAGRSTLKDFI--PVQG-- 74
Db     52 TRVDPQVAVRVRVDGARVVLDDSLVYLALNKPRGMHSTMSDDRGRPCIGDLIERKVRGTKK 111

Qy     74 VYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTL 133
Db    112 LFHVGRLDADTEGLMLLTNDGELAHRLMHPSHEVPKTYLATVTGSVPRGLGRTLRAGIEL 171

Qy    134 NDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWL-KITLYEGRNRQVRRMTAHVGFP 192
Db    172 DDGPAFVDDFAVVD-----------------AIPGKTLVRVTLHEGRNRIVRRLLAAAGFP 215

Qy    193 TLRLIRYAMGSYTLDSLANGEWRDVTPKE 221
Db    216 VEALVRTDIGAVSLGKQRPGSVRALRSNE 244
```

RESULT 12
T36871
probable pseudouridine synthase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21617
A;Accession: T36871

```
Qy    156 NPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGEWR 215
Db    169 ---LDKKKQ--TSIQLTIHEGRNRQVRRMFEAIGHEVKLKREYAFLNLRGLHTGDAR 223

Qy    216 DV 217
Db    224 EL 225
```

RESULT 9
H64874
probable pseudouridylate synthase yciL - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C;Accession: H64874
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cc
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: H64874
A;Molecule type: DNA
A;Status: nucleic acid sequence not shown; translation not shown
A;Residues: 1-291 <BLAT>
A;Cross-references: GB:AE000225; GB:U00096; NID:g1787523; PIDN:AAC74351.1; PID:g1787524;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yciL
C;Superfamily: conserved hypothetical protein HI1243

```
Query Match             21.0%;    Score 245;   DB 2;   Length 291;
Best Local Similarity   34.9%;    Pred. No. 1.8e-14;
Matches 74;  Conservative  23;  Mismatches  79;  Indels  36;  Gaps  7;
```

```
Qy     22 RVE------RFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKDFIP-VQG 73
Db     41 RVEVTPGLKIRIDGHLISVRESAEQICRVLAYYKPEGELCTRNDPEGRPTVFDRLPKLRG 100

Qy     74 V--YAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGV 131
Db    101 ARWIAVGRLDVNTCGLLLFTTDGELANRLMHPSREVEREYAVRVFGQVDDAKLRDLSRGV 160

Qy    132 TLNDGP----TLP-AGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMT 186
Db    161 QLEDGPAAFKTIKFSGGEGINQ-----------------WYNVTLTEGRNREVRRLW 200

Qy    187 AHVGFPTLRLIRYAMGSYTL-DSLANGEWRDV 217
Db    201 EAVGVQVSRLIRVRYGDIPLPKGLPRGGWTEL 232
```

RESULT 10
A64169
hypothetical protein HI1199 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C;Accession: A64169
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64169; MUID:195350630
A;Accession: A64169
A;Molecule type: DNA
A;Status: nucleic acid sequence not shown; translation not shown
A;Residues: 1-357 <TIGR>
A;Cross-references: GB:U32799; GB:L42023; NID:g1574122; PIDN:AAC22853.1; PID:g1574128; T
A;Note: best homolog was a hypothetical protein from Buchnera aphidicola
```

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-371 <MUR>
A;Cross-references: EMBL:AL109848; PIDN:CAB52832.1; GSPDB:GN00070; SCOEDB:SCI51.08c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCI51.08c

```
Query Match             18.5%;  Score 215.5; DB 2; Length 371;
Best Local Similarity   31.2%;  Pred. No. 1e-11;
Matches    65;  Conservative    27;  Mismatches    83;  Indels    33;  Gaps    6;

Qy     30 QATRRTPEPQPTRV----------ILFNKPYDVLPQFTDEAGRSTLKDFIPVQ--GVY 75
             |  || | ||    :|        |:| | |: | |:| || || :: |::|   |: |
Db    171 QGRRVDPEKDEVKVDGLTVATQSYQFFSLNKPAGVVSTMEDPEGRQCLGDYVTNRETRLF 230

Qy     76 AAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYQVEGEPDDASLAK-LRNGVTLN 134
             ||||||||||||||||| | | ||||||||||||||||| |  ||:| | |:||| |:|
Db    231 HVGRLDTETEGVILLTNHGELAHRLTHPRYGVKKTYLAHIVG-PIPRDLGKRLKDGIQLE 289

Qy    135 DGPTLPAGIERVNEPEPLWPRNPPIRERKSIPTSWL-KITLYEGRNRQVRRMTAHVGFPT 193
             |||||||||||||||:|   |:: ||| |:| ||||| ::||||| | |||||| | ||||
Db    290 DG------YARADHFRVVEQTGKNVLVEVTLHEGRKHIVRRMLAEAGFPV 333

Qy    194 LRLIRYAMGSYTLDSLANGEWRDVTPKE 221
             ::| ||:|  | | | ||:|  | |:|
Db    334 DNLVRTAFGPITLGDQKSGWLRRLSNTE 361
```

RESULT 13
S72955
hypothetical protein u0247g - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S72955
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid L247.
A;Reference number: S72580
A;Accession: S72955
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-186 <SMI>
A;Cross-references: EMBL:U00021; NID:g467141; PIDN:AAA50926.1; PID:g467162
C;Genetics:
A;Start codon: GTG

```
Query Match             18.4%;  Score 214; DB 2; Length 186;
Best Local Similarity   34.3%;  Pred. No. 5.7e-12;
Matches    58;  Conservative    24;  Mismatches    65;  Indels    22;  Gaps    4;

Qy     44 ILFNKPYDVLPQFTDEAGRSTLKDFIP--VQG---VYAAGRLDRDSEGLLVLTNDGVLQA 98
             ||||||||||||||||||||||||||:  ::|   :| | |||||||||:||||||||||
Db      9 LALNKPRGMYSTMSDDRGRPCVGDLIERRVRGNKKLFHVGRLDADTEGLLILTNDGELAH 68

Qy     99 RLTQPGKRTGKIYYQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEPLWPRNPP 158
             ||||||||||||||| |:|||||||||||:  |:  |  :||||||||||||| | |
Db     69 RLMHPSHEVSKTYLATVKGAVPRGLGKKLSVGLELDDGPAHVDDFAVVD--------- 118

Qy    159 IRERKSIPTSWL-KITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTL 206
             || |:| || | ::||||| | ||||||| | ||:|||||:||| | |
Db    118 ----AIPGKTLVRLTLHEGRKRIVRRLLTAAGFPVEMLVRTDIGAVSL 161
```

RESULT 14
G71658
hypothetical protein RP544 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 08-Oct-1999
C;Accession: G71658
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U

Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499
A;Accession: G71658
A;Cross-references: EMBL:AL109848; PIDN:CAB52832.1; GSPDB:GN00070; SCOEDB:SCI51.08c
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-235 <AND>
A;Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA14993.1; PID:e134
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: RP544
C;Superfamily: conserved hypothetical protein HI1243

```
Query Match             18.0%;  Score 209.5; DB 2; Length 235;
Best Local Similarity   28.3%;  Pred. No. 2e-11;
Matches    53;  Conservative    39;  Mismatches    70;  Indels    25;  Gaps    5;

Qy     39 QPTRVLFNKPYDVLPQFTDEAGRSTL-KDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQ 97
             | :| | |: | |:.:::|| |  ::||||:|| ::||||||:||||:||||||||||| |
Db     59 QKPRLWIYYKPVGLITTHKDPLSRKTVFEQLIGLPRVISIGRLDLNSEGLLLTNSGDLA 118

Qy     98 ARLTQPGKRTGKIYYQVEGEPDDASLAKLRN---GVTLNDGPTLPAGIERVNEPEPLW 153
             | ::|:||||:||||| |||||||||||||||   :|||||| :|||||||||||||| |
Db    119 HQFEMPASKLKRVYNVRAYGNPNILLKNNVKNLKIDGIFYN-----PHSIKL------- 166

Qy    154 PRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGE 213
             :| ::|| |:| | || | | |||| ::|| | | ||||||::|| |||||||||| |
Db    166 -----LRQNKS--NSWFEVVLFEGKNREIRRIFEYFGLQVNKLIRIQYGALKIGNLKPGN 218

Qy    214 WRDVTPK 220
             :::::||
Db    219 YQEISNK 225
```

RESULT 15
F64987
hypothetical 25.9 kD protein in bcr-rplY intergenic region - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
A;Accession: F64987
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Reference number: A64720; MUID:97426617
A;Accession: F64987
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-231 <BLAT>
A;Cross-references: GB:AE000308; GB:U00096; NID:g1788508; PIDN:AAC75244.1; PID:g17885
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yejD
C;Superfamily: conserved hypothetical protein HI1243

```
Query Match             17.7%;  Score 206; DB 2; Length 231;
Best Local Similarity   33.9%;  Pred. No. 3.9e-11;
Matches    63;  Conservative    31;  Mismatches    66;  Indels    26;  Gaps    7;

Qy     42 RVILFNKPYDVLPQFTDEAGRSTLKDFI--PVQ-GVYAAGRLDRDSEGLLVLTNDGVLQA 98
             :|:|||||||||||||||||||||:| |  :: || | |:|||||||:|||||||||||
Db     61 RYFMLNKPQGVVCS-TDDPDHPTVLYFLDEPVAWKLHAAGRLDIDTTGLVLMTDDGQWSH 119

Qy     99 RLTQPGKRTGKIYYQVEGEPDDASLAKLRNGVTLNDGP--TLPAGIERVNEPEPLWPRN 156
             ||||||||||||||| |||||||||||||| :|||||||  :||||||||||||| | |
Db    120 RITSPRHHCEKTYLVTLESPVADDTAEQFAKGVQLHNEKDLKPAVLEVIT------ 171

Qy    157 PPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDDS-LANGEWR 215
             || |||||| |||||||||||||||||||| | |||||||::|| | |||||  :||||||
Db    171 -------PTQ-VRLTISEGRYHQVKRMFAAVGNHVVELRERIGGITLDADLAPGEYR 220
```

Tue Jun 13 10:30:16 2000

us-09-252-691-7056.rpr

Qy   216 DVTPKE 221
          :|.:|.:|
Db   221 PLTEEE 226

Search completed: June 10, 2000, 11:23:09
Job time: 8313 sec

```
              GenCore version 4.5
         Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:     June 10, 2000, 09:37:11 ; Search time 31.04 Seconds
                                        (without alignments)
                                        217.816 Million cell updates/sec

Title:            US-09-252-691-7056
Perfect score:    1165
Sequence:         1 AIMRQLITPENTMKTSFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters:    83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 08
                 Listing first 45 summaries

Database :  SwissProt_38:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	975	83.7	217	1	YMFC_ECOLI	P75966 escherichia
2	598	51.3	240	1	YMFC_HAEIN	P44827 haemophilus
3	474.5	40.7	261	1	Y612_SYNY3	P72581 synechocyst
4	298	25.6	249	1	Y361_SYNY3	Q55578 synechocyst
5	259	22.2	249	1	YE64_AQUAE	O67444 aquifex aeo
6	253	21.7	229	1	RLUB_BACSU	P35159 bacillus su
7	245	21.0	291	1	YCIL_ECOLI	P37765 escherichia
8	244.5	21.0	357	1	YCIL_HAEIN	P45104 haemophilus
9	233.5	20.0	254	1	YH11_MYCTU	O33210 mycobacteri
10	217	18.7	256	1	YH11_MYCLE	O05668 mycobacteri
11	209.5	18.0	235	1	Y544_RICPR	Q9ZD06 rickettsia
12	206	17.7	231	1	RSUA_ECOLI	P33918 escherichia
13	201.5	17.3	290	1	YJBC_ECOLI	P32684 escherichia
14	200.5	17.2	238	1	Y554_AQUAE	O66829 aquifex aeo
15	198.5	17.0	157	1	YTZF_BACSU	O32068 bacillus su
16	186	16.0	232	1	RSUA_HAEIN	P45124 haemophilus
17	175.5	15.1	258	1	YCIL_BUCAP	P42395 buchnera ap
18	151	13.0	249	1	Y129_BORBU	O51155 borrelia bu
19	119.5	10.3	316	1	YH58_AQUAE	O67638 aquifex aeo
20	113.5	9.7	262	1	YE59_HELPY	P55986 helicobacte
21	111	9.5	253	1	YMDA_CHLAU	Q45826 chloroflexu
22	97	8.3	326	1	RLUD_ECOLI	P33643 escherichia
23	96.5	8.3	327	1	YG29_SYNY3	P74346 synechocyst
24	90.5	7.8	309	1	Y209_MYCPN	P75485 mycoplasma
25	88.5	7.6	302	1	YHCT_BACSU	P54604 bacillus su
26	88	7.6	366	1	QUEA_SYNY3	Q55850 synechocyst
27	86.5	7.4	457	1	IRF7_MOUSE	P70434 mus musculu
28	86.5	7.4	1275	1	RFBC_MYXXA	Q50864 myxococcus
29	85.5	7.3	591	1	RIB2_YEAST	Q12362 saccharomyc
30	85	7.3	376	1	YQCB_ERWCA	Q47417 erwinia car
31	85	7.3	404	1	YG3X_YEAST	P53294 saccharomyc
32	84	7.2	324	1	RLUD_HAEIN	P44445 haemophilus
33	83.5	7.2	2131	1	CINA_DROME	P35500 drosophila
34	83	7.1	485	1	VST2_HEVRH	Q00270 hepatitis e
35	83	7.1	660	1	VST2_HEVMY	Q04611 hepatitis e
36	83	7.1	755	1	AMO_KLEAE	P49250 klebsiella
37	82.5	7.1	406	1	Y496_MYCLE	P54882 mycobacteri
38	81.5	7.0	259	1	STRO_STRGR	P29785 streptomyce
39	81.5	7.0	621	1	CRTI_CERNC	P48537 cercospora
40	80.5	6.9	308	1	Y209_MYCGE	P47451 mycoplasma
41	80.5	6.9	436	1	IF4B_YEAST	P34167 saccharomyc
42	80	6.9	372	1	HMEN_BOMMO	P27609 bombyx mori
43	80	6.9	388	1	ARP2_ACACA	P53487 acanthamoeb
44	79	6.8	245	1	FGFD_HUMAN	O02913 homo sapien
45	79	6.8	245	1	FGFD_MOUSE	P70377 mus musculu

ALIGNMENTS

```
RESULT 1
YMFC_ECOLI
ID   YMFC_ECOLI     STANDARD;      PRT;    217 AA.
AC   P75966;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 24.9 KD PROTEIN IN TRMU-ICDA INTERGENIC REGION.
GN   YMFC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97061202.
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC       STRONG, TO H.INFLUENZAE HI0694.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE000213; AAC74219.1; ALT_INIT.
DR   EMBL; D90748; BAA35957.1; -.
DR   EMBL; D90748; BAA35956.1; -.
DR   ECOGENE; EG13447; YMFC.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   217 AA;   24880 MW;   F7C7A7CEDC5FD3F6 CRC64;

Query Match             83.7%;    Score 975;   DB 1;   Length 217;
Best Local Similarity   85.6%;             Pred. No. 3.8e-79;
```

```
Matches ~185; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

Qy    3 MRQLITPENTMKTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGR 62
          ||||| ||||||||||  | |||||: ||  |:|| |||||||||||||||||||||||
Db    1 MRQFIISENTMQKTSFRNHQVKRFSSQRSTRRKPENQPTRVLFNKPYDVLPQFTDEAGR 60

Qy   63 STLKDFIPVQGVYAAGRLRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDA 122
          .||| ||||||||||||||||||||||||| | ||||||||||||||||||||||   :
Db   61 KTLKEFIPVQGVYAAGRLRDSEGLLVLTNNGALQARLTQPGKRTGKIYYVQVEGIPTQD 120

Qy  123 SLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQV 182
          :| |||||||||||||||||| | |:|| ||||||||||||||||||||||||||||||
Db  121 ALEALRNGVTLNDGPTLPAGAELVDEPAWLMPRNPPIRERKSIPTSWLKITLYEGRNRQV 180

Qy  183 RRMTAHVGFPTLRLIRYAMGSYTLDSLANGEWRDVT 218
          |||||||||||||||||||| |:|||:|||||||:|
Db  181 RRMTAHVGFPTLRLIRYAMGDYSLDNLANGEWREVT 216
```

```
RESULT   2
YMFC_HAEIN
ID   YMFC_HAEIN   STANDARD;      PRT;   240 AA.
AC   P44827;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0694.
GN   HI0694.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd."
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC       STRONG, TO E.COLI YMFC.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U32752; AAC22354.1; -.
DR   TIGR; HI0694.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 240 AA; 27464 MW; 0AD99EC61F52C01C CRC64;
```

```
Query Match
Best Local Similarity    51.3%;    Score 598;  DB 1;  Length 240;
                         59.2%;    Pred. No. 7.8e-46;
Matches 119; Conservative 29; Mismatches 49; Indels 4; Gaps 2;

Qy   15 KTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKDFIPVQGV 74
          |:| || |||||||:|| || |||||||||||||:|||||||||| ||||||||| :
Db   28 KSAVNFHRTFKPKLKPKTLSLDE---TKVVLFNKPFDVLPQFTDEQGRATLKDFISIPNV 84
```

```
Qy   75 YAAGRLRDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLN 134
          |||||||||||||||:| :|| :| | |  |:|||  | |||||||:||   :|||:| |:
Db   85 YAAGRLRDRDSEGLLILNNGLQHRLADPKFKTEKTYWQVEGIPEETDLAQLRKGVELK 144

Qy  135 DGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTL 194
          ||:|| |:| |||:|| |||||||||||:|||||| |:|| :||||||||||||||||||
Db  145 DGVTKSAKVRLISEPN-LWERNPPIRERKNIPTSWLEIKISEGRNRQVRRMTAHIGFPTL 203

Qy  195 RLIRYAMGSYTLDSLANGEWR 215
          ||:| :||  :| ::| |:||
Db  204 RLVRVSMGLLSINGLENGSFR 224
```

```
RESULT   3
Y612_SYNY3
ID   Y612_SYNY3   STANDARD;      PRT;   261 AA.
AC   P72581;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 21.0 KD PROTEIN SLR0612.
GN   SLR0612.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97061201.
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Balroch A.;
RL   Unpublished observations (SEP-1998).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC       INTRODUCED IN POSITION 20 TO PRODUCE THIS ORF.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D90899; BAA16580.1; ALT_FRAME.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 261 AA; 29561 MW; 601A453085C04A69 CRC54;
```

```
Query Match
Best Local Similarity    40.7%;    Score 474.5;  DB 1;  Length 261;
                         50.5%;    Pred. No. 6.8e-35;
Matches 106; Conservative 23; Mismatches 56; Indels 25; Gaps 4;

Qy   31 ATRRTPEPQPTRVILFNKPYDVLPQFTDEAG--RSTLKDFIPVQGVYAAGRLRDSEGLL 88
          | :::||  :|| :|| |||:|| |||| :     :||||:| |  | :| ||||:||||
Db   34 ALNKTPQ----TIVFVKPYGVLCQFTDNSAHPRPTLKDYINLPDLYPVGRLDQDSEGLL 88

Qy   89 VLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNE 148
          :|| .|:|| || |  | :  || :||||   |: | :|| |:|     ||||||: :|
Db   89 LLTSNGLQHRLAHREFAHQRTYFAQVEGSPTDEDLEPLRRGITFADYPTRPAIAKIITE 148

Qy  149 PEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRY------- 200
          |::| |||||| | :||||||| :|:|||||||||||||| |||||||| |:
Db  149 PDFP-PRNPPIYRASIPTSWLSITLTEGRNRQVRRMTAAVGFPTLRLVRVQIVTGRSP 207
```

```
Qy  200 --------AMGSYTLDSLANGEWRDVTPKE 221
             || :| |: | :|| :|| |
Db  208 QQGKGKSAATWCLTIEGLSPGQWRPLTPWE 237
```

RESULT 4
Y361_SYNY3
ID   Y361_SYNY3   STANDARD;   PRT;   249 AA.
AC   Q55578;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 28.2 KD PROTEIN SLR0361.
GN   SLR0361.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96127529.
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D63999; BAA10082.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF01479; S4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   249 AA;   28228 MW;   315F208030180326 CRC64;

Query Match   25.6%;   Score 298;   DB 1;   Length 249;
Best Local Similarity   38.5%;   Pred. No. 2.4e-19;
Matches   72;   Conservative   28;   Mismatches   63;   Indels   24;   Gaps   3;

```
Qy   44 ILFNKPYDVLPQFTDEAGRSTLKDFTPV-----QGVYAAGRLDRDSEGLLVLTNDGVLQA 98
         :||| ||| |||:: | | |:| | ::|     ||| ||:| :|||| |:|||| ||||
Db   67 LLVNKPRDVLSTCDDPRGRKTVLDLLPQDLQRGKGLHPVGRLDRNSTGALLLTNDGELTL 126

Qy   99 RLTQPGKRTGKIYYVOVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLMPRNPP 158
         ||| | | ||:||||: :|| :|| | | |: | ||:||| || :|: :||
Db  127 RLIHPRYHLPKTYDVWLEGNPSDEDLEKWRSGMMLDGKKILPATLEVISENK-------- 179

Qy  159 IRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDS---LANGEW 214
         :|| || |:|:||||:|::| | ||| :||::::: | :|::
Db  179 ------DQIHLLVILTEGRNRQIRRLAEELGLTVLKLHRRTIGPLQLHTRGKVLGSGQF 231

Qy  215 RDVTPKE 221
         | ::|:
Db  232 RFLSPAE 238
```

RESULT 5
YE64_AQUAE
ID   YE64_AQUAE   STANDARD;   PRT;   249 AA.
AC   O67444;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL PROTEIN AQ_1464.
GN   AQ_1464.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AE000741; AAC07410.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF01479; S4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   249 AA;   28518 MW;   29C7A2E7916FDE7C CRC64;

Query Match   22.2%;   Score 259;   DB 1;   Length 249;
Best Local Similarity   34.6%;   Pred. No. 6.6e-16;
Matches   65;   Conservative   28;   Mismatches   77;   Indels   18;   Gaps   3;

```
Qy   37 EPQPTRVLFNKPYDVLPQF-TDEAGRSTLKDFIP--VQGVYAAGRLDRDSEGLLVLTND 93
         :||| |||| |||:|| | :| | |:|| |: |  |||| :||||| | | |||||
Db   59 KPQRKRYIILNKPCCYLTQLGRSPDGRKTIEELIKDIPERVFPVGRLDYNTEGLLILTND 118

Qy   94 GVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLW 153
         | | :| | ||:|| | ||: | |: :|| :|| | :|| ||| || :|:::||::
Db  119 GELANRILHPRYKLPKVYLALVEGKVDQKTLKRMKQGIELEDGFAKPDNIRIV------- 172

Qy  154 PRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGE 213
         :| | ::| :||| ||:| |::||:|:| | | ||
Db  172 -------RYEGKNTLLEITFHBGRKHLVKRFLGAFGHKVKRLKRIAIGPIKLGKLSPGK 223

Qy  214 WRDVTPKE 221
         ||::
Db  224 WRELNQGE 231
```

RESULT 6
RLUB_BACSU
ID   RLUB_BACSU   STANDARD;   PRT;   229 AA.
AC   P35159;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70)
DE   (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE).
GN   RLUB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE; 95020538.
RA   Sorokin A.V., Zumstein E., Azevedo V., Ehrlich S.D., Serror P.;
RT   "The organization of the Bacillus subtilis 168 chromosome region
RT   between the spoVA and serA genetic loci, based on sequence data.";
RL   Mol. Microbiol. 10:385-395(1993).
RN   [2]
RP   CHARACTERIZATION.

RX   MEDLINE: 98325071.
RA   Niu L., Ofengand J.;
RL   Unpublished results, cited by:
RL   Conrad J., Sun D., Englund N., Ofengand J.;
RL   J. Biol. Chem. 273:18562-18566(1998).
CC   -!- FUNCTION: RESPONSIBLE FOR SYNTHESIS OF PSEUDOURIDINE FROM URACIL-
CC   2650 IN 23S RIBOSOMAL RNA.
CC   -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE
CC   5'-PHOSPHATE + H(2)O.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L09228; AAA67493.1; -.
DR   EMBL; Z99116; CAB14248.1; -.
DR   PIR; S45555; S45555.
DR   SUBTILIST; BG10530; RLUB
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF00849; YABO; 1.
DR   PFAM; PF01479; S4; 1.
KW   Lyase.
SQ   SEQUENCE   229 AA; 26025 MW; 0CD54C2E7A5ADE8 CRC64;

Query Match   21.7%; Score 253; DB 1; Length 229;
Best Local Similarity   35.2%; Pred. No. 2e-15;
Matches 64; Conservative 31; Mismatches 69; Indels 18; Gaps 5;

Qy   39 QPTRVILFNKPYDVLPQFTDEAGRSTLKDF---IPVQGVYAAGRLDRDSEGLLVLTNDGV 95
Db   59 EPVYFLLY-KPRGVISAAQDDKGRKVVTDFFKNIP-QRIYPIGRLDYDTSGLLLLTNDGE 116

Qy   96 LQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPR 155
Db 117 FANKLMHPKYEIDKTYVAVKGIPPKELLRKLERGIRLEEGKTAPAKAKLLS-------- 169

Qy 156 NPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGEWR 215
Db 169 ---LDKKKQ--TSIIQLTIHEGRNRQVRRMFEAIGHEVIKLKREEYAFLNLRGLHTGDAR 223

Qy 216 DV 217
Db 224 EL 225

RESULT 7
YCIL_ECOLI
ID   YCIL_ECOLI   STANDARD;   PRT;   291 AA.
AC   P37765;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 32.7 KD PROTEIN IN TRPL-BTUR INTERGENIC REGION (ORF4).
GN   YCIL.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE: 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).

RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE: 97251357.
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA   Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA   Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,
RA   Oshima T., Saito N., Sampei G., Seki Y., Sivasundaram S.,
RA   Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA   Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
RN   [3]
RP   SEQUENCE OF 1-243 FROM N.A.
RC   STRAIN=K12 / W3110;
RA   Milkman R., McKane M.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 217-291 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE: 89123012.
RA   Lundrigan M.D., Kadner R.J.;
RT   "Altered cobalamin metabolism in Escherichia coli btuR mutants
RT   affects btuB gene regulation.";
RL   J. Bacteriol. 171:154-161(1989).
RN   [5]
RP   IDENTIFICATION.
RX   MEDLINE: 95075659.
RA   Borodovsky M., Rudd K.E., Koonin E.V.;
RT   "Intrinsic and extrinsic approaches for detecting genes in a
RT   bacterial genome.";
RL   Nucleic Acids Res. 22:4756-4767(1994).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   STRONG, TO H.INFLUENZAE HI1199.
CC   -!- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC   FRAMESHIFT.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is not removed. Usage by and for commercial
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U18111; AAB59990.1; -.
DR   EMBL; AE000225; AAC74351.1; -.
DR   EMBL; D90764; CAB20715.1; -.
DR   EMBL; D90765; CAB20740.1; -.
DR   EMBL; M21528; -; NOT_ANNOTATED_CDS.
DR   ECOGENE; EG12433; YCIL.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF01479; S4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   291 AA; 32711 MW; 101DF8FD55CB557D CRC64;

Query Match   21.0%; Score 245; DB 1; Length 291;
Best Local Similarity   34.9%; Pred. No. 1.4e-14;
Matches 74; Conservative 23; Mismatches 79; Indels 36; Gaps 7;

Qy   22 RVE------RFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKDFIP-VQG 73
Db   41 RVEVTPGLKIRIDGHLISVRESAEQICRVLAYYKPEGELCTRNDPEGRPTVFDRLPKLRG 100

Qy   74 V--YAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGV 131
Db 101 ARWIAVGRLDVNTCGLLFTTDGELANRLMHPSREVEREYAVRVFGQVDDAKLRDLSRGV 160

Qy 132 TLNDGP----TLP-AGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMT 186

Db 161 QLEDGPRAFKTIKFSGGEGINQ------------------WYNVLTEGRNREVRRLW 200

Qy 187 AHVGFPTLRLIRYAMGSYTL-DSLANGEWRDV 217
            || |||||||| |::   : : || ::
Db 201 EAVGVQVSRLIRVRYGDIPLPKGLPRGGWTEL 232

RESULT 8
YCIL_HAEIN
ID YCIL_HAEIN   STANDARD;   PRT;   357 AA.
AC P45104;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN HI1199.
GN HI1199.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   STRONG, TO E.COLI YCIL.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; U32799; AAC22853.1; -.
DR TIGR; HI1199; -.
DR PROSITE; PS01149; PSI_RSU; 1.
DR PFAM; PF01479; S4; 1.
KW Hypothetical protein.
SQ SEQUENCE 357 AA; 40636 MW; 0A31F0D384CBBE82 CRC64;

Query Match   21.0%; Score 244.5; DB 1; Length 357;
Best Local Similarity 37.2%; Pred. No. 2e-14;
Matches 67; Conservative 24; Mismatches 70; Indels 19; Gaps 5;

Qy   42 RVILFNKPYDVLPQFTDEAGRSTLKDFIP-VQGV--YAAGRLDRDSEGLLVLTNDGVLQA 98
            | | :: || | |::|: : : | : : | | |||| :: |||| | | |
Db 141 RVLMYYKPEGELCTRSDPEGRATVFDRLPRLTGSRWIAVGRLDINTSGLLLFTTDGELAN 200

Qy   99 RLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPP 158
            ||| | : | ||||||||||||||| ||||| |||:||||||||||| ||||
Db 201 RLMHPSREVEREYSVRVFGQVDDAMLARLRKGVQLEDG------LANFKEIKFTGGVG 252

Qy 159 IRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTL-DSLANGEWRDV 217
            | : | : || ||||||:|||:::|| | || ||||| :|| : ::
Db 253 INQ------WIDVTLMEGRNREVRRLWESQIQVSRLIRIRYGNIKLMKGLPRGGWEEM 305

RESULT 9
YH11_MYCTU

ID YH11_MYCTU   STANDARD;   PRT;   254 AA.
AC O33210;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 27.6 KD PROTEIN RV1711.
GN RV1711 OR MTCI125.33.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE; 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; Z98268; CAB10968.1; -.
DR PROSITE; PS01149; PSI_RSU; 1.
DR PFAM; PF00849; YABO; 1.
DR PFAM; PF01479; S4; 1.
DR TUBERCULIST; Rv1711; -.
KW Hypothetical protein.
SQ SEQUENCE 254 AA; 27597 MW; EB2D95D7CF0FDF26 CRC64;

Query Match   20.0%; Score 233.5; DB 1; Length 254;
Best Local Similarity 32.5%; Pred. No. 1.2e-13;
Matches 68; Conservative 25; Mismatches 81; Indels 35; Gaps 5;

Qy   32 TRRTPEPQPTRV--------------ILFNKPYDVLPQFTDEAGRSTLKDFI--PVQG--- 74
            | | | : | |:| | || | | ||| |: : |: :
Db   52 TRVDPQVAVVRVRVDGARVLDDSLVYLALNKPRGMHSTMSDDRGRPCIGDLIERKVRGTKK 111

Qy   74 VYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTL 133
            :: |||| |:|||:|||| ||| || ||| | |||:: | :|
Db 112 LFHVGRLDADTEGLMLLTNDGELAHRLMHPSHEVPKTYLATVTGSVPRGLGRTLRAGIEL 171

Qy 134 NDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWL-KITLYEGRNRQVRRMTAHVGFP 192
            |:   : :: || | ||: | :|::||||| : ||:||||:|||| | |
Db 172 DDGPAFVDDFAVVD------------------AIPGKTLVRVTLHEGRNRIVRRLLAAAGFP 215

Qy 193 TLRLIRYAMGSYTLDSLANGEWRDVTPKE 221
            :: || :|| | :| ::| :
Db 216 VEALVRTDIGAVSLGKQRPGSVRALRSNE 244

RESULT 10
YH11_MYCLE
ID YH11_MYCLE   STANDARD;   PRT;   256 AA.
AC O05668; Q49886;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 28.1 KD PROTEIN CB1351.03C.

```
GN   MLCB1351.03C OR U0247G.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murphy L., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 71-256 FROM N.A.
RA   Smith D.R., Robison K.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95117; CAB08276.1; -.
DR   EMBL; U00021; AAA50926.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF00849; YABO; 1.
DR   PFAM; PF01479; S4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   256 AA;  28113 MW;  686746EF490FB695 CRC64;

Query Match             18.7%;  Score 217.5;  DB 1;  Length 256;
Best Local Similarity   32.5%;  Pred. No. 3.1e-12;
Matches   63;  Conservative   25;  Mismatches   71;  Indels   35;  Gaps    5;

Qy    32 TRRTPEPQPTRV----------ILFNKPYDVLPQFTDEAGRSTLKDFIP--VQG--- 74
Db    54 TRVDPDVSVRVDGVKVVVDDSLVYLALNKPRGMYSTMSDDRGRPCVGDLIERRVRGNKK 113

Qy    74 VYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTL 133
Db   114 LFHVGRLDADTEGLLLLTNDGELAHRLMHPSHEVSKTYLATVKGAVPRGLGKKLSVGLEL 173

Qy   134 NDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWL-KITLYEGRNRQVRRMTAHVGFP 192
Db   174 DDGPAHVDDFAVVD-----------AIPGKTIVRLLTHEGRKRIVRRLLTAAGFP 217

Qy   193 TLRLIRYAMGSYTL 206
Db   218 VEMLVRTDIGAVSL 231

RESULT 11
Y544_RICPR
ID   Y544_RICPR    STANDARD;      PRT;   235 AA.
AC   Q92D06;
DT   15-FEB-2000 (Rel. 39, Created)
DT   15-FEB-2000 (Rel. 39, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN RP544.
GN   RP544.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 99039499.
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235272; CAA14993.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   235 AA;  26815 MW;  9E09C0BAC46938A7 CRC64;

Query Match             18.0%;  Score 209.5;  DB 1;  Length 235;
Best Local Similarity   28.3%;  Pred. No. 1.4e-11;
Matches   53;  Conservative   39;  Mismatches   70;  Indels   25;  Gaps    5;

Qy    39 QPTRVILFNKPYDVLPQFTDEAGRSTL-KDFIPVQGVVAAGRLDRDSEGLLVLTNDGVLQ 97
Db    59 QKPRLWIYYKPVGLITTHKDPLSRKTVFEQLIGLPRVISIGRLDNSEGLLLLTNSGDLA 118

Qy    98 ARLTQPGKRTGKIYYVQVEGEPDDASLAKLRN---GVTLNDGPTLPAGIERVNEPEWLW 153
Db   119 HQFEMPASKLKRVINVRAYGNPNILLKNNYKNLKIDGIFYN-----PHSIKL------ 166

Qy   154 PRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGPPTLRLIRYAMGSYTLDSLANGE 213
Db   166 ----LRQNKS--NSWFEVVLFEGKNREIRRIFEYFGLQVNKLIRIQYGALKIGNLKPGN 218

Qy   214 WRDVTPK 220
Db   219 YQEISNK 225

RESULT 12
RSUA_ECOLI
ID   RSUA_ECOLI    STANDARD;      PRT;   231 AA.
AC   P33918;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RIBOSOMAL SMALL SUBUNIT PSEUDOURIDINE SYNTHASE A (EC 4.2.1.70) (16S
DE   PSEUDOURIDYLATE 516 SYNTHASE) (16S PSEUDOURIDINE 516 SYNTHASE) (URACIL
DE   HYDROLYASE).
GN   RSUA.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 1-13, AND CHARACTERIZATION.
RX   MEDLINE; 95337112.
RA   Wrzesinski J., Bakin A., Nurse K., Lane B.G., Ofengand J.;
RT   "Purification, cloning, and properties of the 16S RNA pseudouridine
RT   516 synthase from Escherichia coli.";
RL   Biochemistry 34:8904-8913(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / BHB2600;
RA   Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA   Church G.M.;
RL   Submitted (OCT-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
```

RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: FORMATION OF PSEUDOURIDINE FROM URACIL-516 IN
CC   16S RIBOSOMAL RNA.
CC   -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE
CC   5'-PHOSPHATE + H(2)O.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---
DR   EMBL; U00008; AAA16377.1; -.
DR   EMBL; AE000308; AAC75244.1; -.
DR   ECOGENE; EG12044; RSUA.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Lyase.
KW   PFAM; PF01479; S4; 1.
SQ   SEQUENCE   231 AA;   25865 MW;   04EB7154931C58F7 CRC64;

Query Match   17.7%;   Score 206;   DB 1;   Length 231;
Best Local Similarity 33.9%;   Pred. No. 2.8e-11;
Matches   63;   Conservative   31;   Mismatches   66;   Indels   26;   Gaps   7;

Qy   42 RVILFNKPYDVLPQFTDEAGRSTLKDFI--PVQ-GVYAAGRLDRDSEGLLVLTNDGVLQA 98
               :::: ::::::::::::: : :::::::: ::::::::::::::: :
Db   61 RYFMLNKPQGVVCS-TDDPDHPTVLYFLDEPVAWKLHAAGRLDIDTTGLVLMTDDGQWSH 119

Qy   99 RLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGP--TLPAGIERVNEPEWLWPRN 156
               :: : : : :::: :: ::::::::::::::: :::: : : ::::::::
Db   120 RITSPRHHCEKTYLVTLESPVADDTAEQFAKGVQLHNEKDLTKPAVLEVIT------- 171

Qy   157 PPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDS-LANGEWR 215
               || :::: || |||||| | : : || || ||::| || ||
Db   171 -PTQ-VRLTISEGRYHQVKRMFAAVGNHVELHRERIGGITLDADLAPGEYR 220

Qy   216 DVTPKE 221
               ||::|
Db   221 PLTEEE 226

RESULT 13
YJBC_ECOLI
ID   YJBC_ECOLI   STANDARD;   PRT;   290 AA.
AC   P32684;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 32.5 KD PROTEIN IN PEPE-LYSC INTERGENIC REGION.
GN   YJBC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 94089392.
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---

DR   EMBL; U00006; AAC43116.1; -.
DR   EMBL; AE000475; AAC76992.1; -.
DR   ECOGENE; EG11921; YJBC.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF01479; S4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   290 AA;   32476 MW;   C88C830CBE5DF2FA CRC64;

Query Match   17.3%;   Score 201.5;   DB 1;   Length 290;
Best Local Similarity 29.6%;   Pred. No. 9.5e-11;
Matches   55;   Conservative   34;   Mismatches   76;   Indels   21;   Gaps   4;

Qy   37 EPQPTRVILFNKPYDVLPQFTDEAGRSTLKDFI-PVQGVYAAGRLDRDSEGLLVLTNDGV 95
               | : :|||:: ::::::::::::::::: |:: : ||||||||||||::|||||
Db   63 EAEDLVLIALNKPVGIV-STTEDGERDNIVDFVNHSKRVFPIGRLDKDSQGLLFLTNHGD 121

Qy   96 LQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPR 155
               | :| | : : || | :: : || :
Db   122 LVNKILRAGNDHEKEYLVTVDKPITEEFIRGMSAGVPI-------------LGTV 163

Qy   156 NPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGEWR 215
               : :: :::: :::| :|:|||| || : : : :: : : : | :: ||||
Db   164 TKKCKVKKEAPFVF-RITLVQGLNRQIRRMCEHFGYEVKKLERTRIMNVSLSGIPLGEWR 222

Qy   216 DVTPKE 221
               |::|
Db   223 DLTDDE 228

RESULT 14
Y554_AQUAE
ID   Y554_AQUAE   STANDARD;   PRT;   238 AA.
AC   O66829;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL PROTEIN AQ_554.
GN   AQ_554.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE; 98196666.
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   ---
DR   EMBL; AE000695; AAC06794.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF01479; S4; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   238 AA;   27862 MW;   904DDD431CD3E870 CRC64;

Query Match   17.2%;   Score 200.5;   DB 1;   Length 238;

Job time: 6514 sec

```
Best Local Similarity   27.4%;   Pred. No. 9e-11;
Matches   58;   Conservative   43;   Mismatches   76;   Indels   35;   Gaps   5;

Qy    20  KHRVERFSSRQATRRTPEPQPTRVILFNKPYDVL--------PQFTDEAGRSTLKDFIP  70
              ::|:.    :    ::|:.   ||      |       :::    |
Db    38  EYRVKEGEEVEVEGKSVKPKKNVYLMLYKPKGYLSTEEDKKYPSFLE----LIREHFP  92

Qy    71  VQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNG  130
              : :::||||| |:||||::|||   ||||||: |||||  : :||: : | ||
Db    93  SRKLFSAGRLDVDAEGLLLITDDGELAHRLTHPKWKVEKEYIVRLDRDIGDEELKKLYE-  152

Qy   131  VTLNDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVG  190
              | |:|   :|:::     |    ||| |:  ||||   | ||
Db   152  VKLEEKPVQLVKAEKLS------------------GDTVKAILTEGRHHVVKRLFKAVG  192

Qy   191  FPTLRLIRYAMGSYTLD-SLANGEWRDVTPKE  221
              :| :|| || ::  |||:|:::| :|
Db   193  HNVVYLKRTRVGNLRLDENMEPGEWRELTEEE  224


RESULT 15
YTZF_BACSU
ID   YTZF_BACSU       STANDARD;       PRT;      157 AA.
AC   O32068;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 17.7 KD PROTEIN IN AMYX-OPUD INTERGENIC REGION.
GN   YTZF.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   -------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in  no  way
CC   modified and this statement is not removed. Usage  by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------
DR   EMBL; Z99119; CAB14981.1; -.
DR   SUBTILIST; BG13940; YTZF.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE  157 AA;  17695 MW;  D378BFA4E89DCE1 CRC64;

Query Match        17.0%;   Score 198.5;   DB 1;   Length 157;
Best Local Similarity   36.2%;   Pred. No. 7.9e-11;
Matches   54;   Conservative   25;   Mismatches   53;   Indels   17;   Gaps   4;

Qy    75  YAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLN  134
              : |||||:|:|| |:||||| |:|||||   |    | | :: :||:  ::  | |
Db    15  FPAGRLDKDTEGFLLLTNDGQLAHRLLSPKKHVPKTYEVHLKSQISREDISDLETGVVIE  74

Qy   135  DG-PTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPT  193
              |  | |||:: :::|     ::: :|: ::| ::  |:   || ::|
Db    75  GGYKTKPAKAE---------IKTNDSGNTV-IYLITEGKYHQVKQMAKAVGNEV  119

Qy   194  LRLIRYAMGSYTLD-SLANGEWRDVTPKE  221
              :|  || :||  ::|| :||:|::|:| :|
Db   120  VYLKRLSMGRVSLDPALAPGEYRELTEEE  148
```

Search completed: June 10, 2000, 11:25:45

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     June 10, 2000, 09:36:29 ; Search time 53.1 Seconds
                                       (without alignments)
                                       289.871 Million cell updates/sec

Title:            US-09-252-691-7056
Perfect score:    1165
Sequence:         1 AIMRQLITPENTMKTSFRK..........SYTLDSLANGEWRDVTPKEN 222

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters:    225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database :    SPTREMBL_12:*
              1:  sp_archea:*
              2:  sp_bacteria:*
              3:  sp_fungi:*
              4:  sp_human:*
              5:  sp_invertebrate:*
              6:  sp_mammal:*
              7:  sp_mhc:*
              8:  sp_organelle:*
              9:  sp_phage:*
             10:  sp_plant:*
             11:  sp_rodent:*
             12:  sp_virus:*
             13:  sp_vertebrate:*
             14:  sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	264.5	22.7	239	2	Q9WYA2	Q9wya2 thermotoga
2	209.5	18.0	235	2	Q9ZD06	Q9zd06 rickettsia
3	178.5	15.3	235	2	Q9Z742	Q9z742 chlamydia p
4	176	15.1	338	5	O96270	O96270 plasmodium
5	153.5	13.2	241	2	O84728	O84728 chlamydia t
6	120	10.3	261	2	O83472	O83472 treponema p
7	112.5	9.7	262	2	Q9ZJG0	Q9zjg0 helicobacte
8	112.5	9.7	304	2	Q9X038	Q9x038 thermotoga
9	107.5	9.2	1464	5	Q9XZY5	Q9xzy5 leishmania
10	95.5	8.2	217	2	O50765	O50765 mycobacteri
11	88.5	7.6	454	1	O54664	O54664 halobacteri
12	88	7.6	303	2	O84108	O84108 chlamydia t
13	86.5	7.4	454	5	Q9XZQ7	Q9xzq7 caenorhabdi
14	84.5	7.3	473	2	Q9Z664	Q9z664 zymomonas m
15	84.5	7.3	660	5	Q24717	Q24717 drosophila
16	84.5	7.3	689	5	Q24715	Q24715 drosophila
17	84.5	7.3	693	5	Q24712	Q24712 drosophila
18	84.5	7.3	701	5	Q24713	Q24713 drosophila
19	84.5	7.3	704	5	Q24716	Q24716 drosophila
20	84.5	7.3	714	5	Q24714	Q24714 drosophila
21	84.5	7.3	946	10	Q42571	Q42571 arabidopsis
22	84.5	7.3	946	10	Q42587	Q42587 arabidopsis
23	84.5	7.3	1020	10	Q37145	Q37145 arabidopsis
24	84.5	7.3	1020	10	Q37146	Q37146 arabidopsis
25	84	7.2	660	12	Q68985	Q68985 hepatitis e
26	84	7.2	2354	5	Q24135	Q24135 drosophila
27	83.5	7.2	2077	5	Q24528	Q24528 drosophila
28	83.5	7.2	2110	5	Q24532	Q24532 drosophila
29	83.5	7.2	2118	5	Q24531	Q24531 drosophila
30	83.5	7.2	2121	5	Q24529	Q24529 drosophila
31	83	7.1	660	12	Q81878	Q81878 hepatitis e
32	82.5	7.1	4151	2	O53490	O53490 mycobacteri
33	82	7.0	235	2	Q49848	Q49848 mycobacteri
34	81.5	7.0	1091	13	Q9YH41	Q9yh41 gallus gall
35	81.5	7.0	4162	13	Q98918	Q98918 gallus gall
36	81	7.0	303	10	O80967	O80967 methanobact
37	81	7.0	419	1	O26396	O26396 methanobact
38	81	7.0	1129	4	O60292	O60292 homo sapien
39	80	6.9	86	10	O80966	O80966 arabidopsis
40	80	6.9	436	12	Q9WI48	Q9wi48 hepatitis e
41	80	6.9	660	12	Q69419	Q69419 hepatitis e
42	80	6.9	1145	10	O04146	O04146 acetabulari
43	79.5	6.8	262	5	O16582	O16582 caenorhabdi
44	79	6.8	660	12	Q69411	Q69411 hepatitis e
45	79	6.8	660	12	Q81871	Q81871 hepatitis e

ALIGNMENTS

```
RESULT     1
Q9WYA2
ID   Q9WYA2       PRELIMINARY;      PRT;   239 AA.
AC   Q9WYA2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   16S PSEUDOURIDYLATE SYNTHASE.
GN   TM0264.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE: 99287316.
RA   NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA   HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA   MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA   STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA   HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA   SMITH H.O., VENTER J.C., FRASER C.M.;
RT   "Evidence for lateral gene transfer between Archaea and bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA   HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA   MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA   STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA   HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA   SMITH H.O., VENTER J.C., FRASER C.M.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001708; AAD35352.1; -.
SQ   SEQUENCE   239 AA;  27160 MW;  511D06DD CRC32;

Query Match             22.7%;  Score 264.5;  DB 2;  Length 239;
Best Local Similarity   35.8%;  Pred. No. 4.3e-16;
Matches   64;  Conservative   31;  Mismatches   65;  Indels   19;  Gaps  3;

Qy     44 ILFNKPYDVLPQFTDEAGRSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQP 103
          |||.||     | : :  |||:|: |||:|::|||| |||||:||||:||||  |:||
Db     62 ILFYKPSGYVTSTKDPHSETIMEFLPPLKGIFPVGRLDKDAEGLLIITNDGDFAHRVISP 121
```

Qy   104 GKRTGKIYYQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPIRERK 163
         ||||||| : :||:|||||| || :|||
Db   122 KWSVPEKEYIVKVEGEVTEDKIEKLKNGVTLRDG--FFAKAKRV----------------E 163

Qy   164 SIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRIRYAMGSYTL-DSLANGEWRDVTPKE 221
         ||| :| || :||||:||| || : | | :| :||| ::|
Db   164 KLSNDTLKIVITEGKYHQIKRMTAAVGLKTVHLKRTRIGGLVLPDDMKPGEYRFLSEEE 222

RESULT 2
ID   Q9ZD06   PRELIMINARY;   PRT;   235 AA.
AC   Q9ZD06;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HYPOTHETICAL 26.8 KD PROTEIN.
GN   RP544.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 99039499.
RA   ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA   SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA   ERIKSSON A.S., WINKLER H.H., KURLAND C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RA   ANDERSSON S.G.E.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ235272; CAA14993.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 235 AA; 26815 MW; E5704F25 CRC32;

Query Match   18.0%; Score 209.5; DB 2; Length 235;
Best Local Similarity   28.3%; Pred. No. 3.2e-11;
Matches 53; Conservative 39; Mismatches 70; Indels 25; Gaps;

Qy   39 QPTRVILFNKPYDVLPQFTDEAGRSTL-KDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQ 97
         : || :|| ||: | || |||| |||||||:||| ||
Db   59 QKPRLWIIYKPVGLITTHKDPLSRKTVFEQLIGLPRVISIGRLDLNSEGLLLLTNSGDLA 118

Qy   98 ARLTQPGKRTGKIYYQVEGEPDDASLAKLRN----GVTLNDGPTLPAGIERVNEPEWLW 153
         | : | :|| ||:|:|:||||||||||:|| :|||||||||||||||||||||||
Db 119 HQFEMPASKLKRVTNVRAYGNPNILLKNNYKNLKIDGIFYN-----PHSIKL-------- 166

Qy 154 PRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLRYAMGSYTLDSLANGE 213
             || || ::||||:|||| :| ::||| | ::: :|
Db 166 -----LRQNKS--NSWFEVVLFEGKNREIRRIFEYFGLQVNKLIRIQYGALKIGNLKPGN 218

Qy 214 WRDVTPK 220
         ::::: |
Db 219 YQEISNK 225

RESULT 3
ID   Q9Z742   PRELIMINARY;   PRT;   235 AA.
AC   Q9Z742;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PREDICTED PSEUDOURIDINE SYNTHASE.
GN   YJBC.

OS   Chlamydia pneumoniae.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RA   KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,
RA   GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;
RT   "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001667; AAD19002.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
SQ   SEQUENCE 235 AA; 25942 MW; 80240286 CRC32;

Query Match   15.3%; Score 178.5; DB 2; Length 235;
Best Local Similarity   27.0%; Pred. No. 1.8e-08;
Matches 40; Conservative 26; Mismatches 63; Indels 19; Gaps 1;

Qy   74 VYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYQVEGEPDDASLAKLRNGVTL 133
         |: |||||| |||||||:||:|| :|| | || |||:||||| ::| |||
Db   98 VFTVGRLDKETSGLLLVTNDGEFANKIIHPSSGITKEYLLKVSRDVSAKDLGKLMEGTFI 157

Qy 134 NDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPT 193
             ||| :|| :| ::|
Db 158 DGKHVRPVSVTKIRR---------------------GTVKIVVSEGKKHEIRLFADAAGFPI 198

Qy 194 LRLIRYAMGSYTLDSLANGEWRDVTPKE 221
         | | : | |||||::|| :||:::|||
Db 199 LELKRIRIGSLVLGGLRYGEYRELTDAE 226

RESULT 4
ID   O96270   PRELIMINARY;   PRT;   338 AA.
AC   O96270;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PSEUDOURIDINE SYNTHETASE (RSUA FAM.), 1ST EUK. MEMBER (OO).
GN   PFB0890C.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99021743.
RA   GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA   KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
RA   SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA   SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA   FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
DR   EMBL; AE001423; AAC71967.1; -.
SQ   SEQUENCE 338 AA; 40297 MW; AFDC23C5 CRC32;

Query Match   15.1%; Score 176; DB 5; Length 338;
Best Local Similarity   27.1%; Pred. No. 5e-08;
Matches 57; Conservative 38; Mismatches 83; Indels 32; Gaps 6;

Qy   44 ILFNKPYDVLPQFTDEAGRSTLKDFIPVQ----GVYAAGRLDRDSEGLLVLTNDGVLQA 98
         :: :|| | | | || || | : || :||||:||||||||||| :::|:
Db 120 IVLHKPKGLLCTSNDEKNRKSIYTLFPEEMLQKYRLVTVGRLDRNTSGVLLLTNDYAWVN 179

Qy   99 RLTQPGKRTGKIYYQVEGEPDDASLAKLRNGVTL-NDGPTLPAGIERVNE---------- 149
         ::|:|| ||:| :|:|| | |:||:||||||:|| :|||:||||||||
Db 180 KLTHPKYQRIRTYRVHIEGPVKMNALKELARGIYLEEDEKTQPKKIYNYKSREKSNIDD 239

Qy 149 ------PEWLWPRNPP----IRERK-----SIPTSWLKITLYEGRNRQVRRMTAHVGFP 192
         || |||:|||||||||| :|||||||||||||||||||||||||
Db 240 KKKKKMSKMKKKTNPAFIEILREEKIIKEDTKKIITVLNISIKEGRNRQIRKMFQQINQP 299

```
Qy   193  TLRLIRYAMGSYTLDSL-ANGEWRDVTPKE 221
               ::|:|:::||:||:::||:|  ||
Db   300  VIKIKRTSFENITLKNIYFPKQYRELNQKE 329

RESULT 5
084728
ID   084728      PRELIMINARY;    PRT;    241 AA.
AC   084728;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PREDICTED PSEUDOURIDINE SYNTHASE.
GN   YJBC.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   Science 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001343; AAC68318.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF01479; S4; 1.
SQ   SEQUENCE 241 AA;  26395 MW;  C6D4B6B0 CRC32;

Query Match             13.2%;  Score 153.5;  DB 2;  Length 241;
Best Local Similarity   25.0%;  Pred. No. 3.1e-06;
Matches   37;  Conservative  27;  Mismatches  65;  Indels  19;  Gaps

Qy    74  VYAAGRLDRDSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTL 133
              ::|||||:::||:::||||  |:|||||||  |:|||||||:  |  |   |||
Db    99  LFTVGRLDKETSGLIIVTNDGEFANRVIHPSFGITKEYLLKVSRDVTARDLETLMAGTVI 158

Qy   134  NDGPTLPAGIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPT 193
               ||  :::|           |:|||||:|:|:|||||||:|:|  :||
Db   159  DGKVVRPVSVKKVRR----------------GTIKIIVNEGKKHEIRLFAEAAGLQL 199

Qy   194  LRLIRYAMGSYTLDSLANGEWRDVTPKE 221
              |   ||   |:||:|:||:|:||:||
Db   200  LELKRIRIGSLVLGGLPYGKYRELTDSE 227

RESULT 6
083472
ID   083472      PRELIMINARY;    PRT;    261 AA.
AC   083472;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   TP0459.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   MEDLINE; 98332770.
RX   MEDLINE; 98332770.
RA   FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA   DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA   SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA   KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T...
```

```
RA   MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJI C., GARLAND S.,
RA   HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA   VENTER J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA   DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA   SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA   KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA   MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJI C., GARLAND S.,
RA   HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA   VENTER J.C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001223; AAC65447.1; -.
DR   TIGR; TP0459; -.
DR   PFAM; PF01479; S4; 1.
SQ   SEQUENCE 261 AA;  29216 MW;  906F6CCB CRC32;

Query Match             10.3%;  Score 120;  DB 2;  Length 261;
Best Local Similarity   24.9%;  Pred. No. 0.0033;
Matches   50;  Conservative  24;  Mismatches  85;  Indels  42;  Gaps  6;

Qy    42  RVILFNKPY----DVLPQFTD------EAGRSTLK-----DFIP---VQGVYAAGRLDR 82
              |:|| |  ||    | |:  |||:   ||||
Db    70  RVLLYKPVGYICSLAPQFPAGYAHTQVRAGPSKQEYARAIDLVQPAYQERLYHIGRLDV 129

Qy    83  DSEGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAG 142
              ||| ||: ||||  || | |:|:||:  |   |  |: ||
Db   130  RSEGALLFTNDGSFAQALGHPRSGIEKEYIVETREPVPAALLSSFVRGV---------- 179

Qy   143  IERVNEPEWLWPRNPPIR--ERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYA 200
              :   |||:   ::| ||:|:|||  :|| :|:|
Db   179  -----------WVEGCRYRCVRARHLAAQCVQLVLVEGKKREIRVVFEAWGQDVRLVRVR 228

Qy   201  MGSYTLDSLANGEWRDVTPKE 221
              :||:|::||::||:||:||
Db   229  IGRVRLGPLRPGAFRTMRLQE 249

RESULT 7
Q92JG0
ID   Q92JG0      PRELIMINARY;    PRT;    262 AA.
AC   Q92JG0;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE.
GN   JHP1352.
OS   Helicobacter pylori J99.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-J99.
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001558; AAD06937.1; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
SQ   SEQUENCE 262 AA;  30191 MW;  848C3398 CRC32;

Query Match              9.7%;  Score 112.5;  DB 2;  Length 262;
```

```
Best Local Similarity  25.4%;  Pred. No. 0.015;
Matches   52;  Conservative  37;  Mismatches  63;  Indels  53;  Gaps  10;

Qy    43 VILFNKPYDVLPQFTDEAGRSTLKD--------FIPVQGVYAAGRLDRDSEGLLVLTNDG  94
         :::::||                          :||||            ||
Db    68 VLVYHKPKGELVSKADPLKRHVIYESLEKKYAHFAPV------GRIDFASEGVLLLSDSK 121

Qy    95 VLQARLTQPGKRTGKIYYVQEG----EPPDASLAKLRNGVTLNDGPTLPAGIERVNEPE 150
         |:||:                     ::|  ||||  :: :|||:
Db   122 AVVSALMHANLE--KEMLVKIQGFVTREMENA----MQEGLKL-ENATKGA--------- 166

Qy   151 WLWPRNPPIRERKSIP------------TSWLKITLYEGRNRQVRRMTA--HVGFPTLRL 196
         :: ||:                      ||:||||||:|:::|| :  :   || ||
Db   166 ---HQKTPIKSMEFAPFIGYELIKNHAKYSKLRVIINEGKNRELRRFFAFFNAGVLDLRR 222

Qy   197 IRYAMGSYTLDSLANGEWRDVTPKE 221
         :|: ||:||:|:|: || : :
Db   223 VRY--GFVNLNALPVGKARFLNRQE 245


RESULT  8
Q9X038
ID   Q9X038       PRELIMINARY;      PRT;    304 AA.
AC   Q9X038;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE C.
GN   TM0940.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99287316.
RA   NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA   HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA   MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA   STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA   HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA   SMITH H.O., VENTER J.C., FRASER C.M.;
RT   "Evidence for lateral gene transfer between Archaea and bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA   HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA   MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA   STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA   HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA   SMITH H.O., VENTER J.C., FRASER C.M.;
RP   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001757; AAD36021.1; -.
DR   PROSITE; PS01129; PSI_RLU; 1.
SQ   SEQUENCE  304 AA;  35750 MW;  771880E6 CRC32;

Query Match           9.7%;  Score 112.5;  DB 2;  Length 304;
Best Local Similarity  26.1%;  Pred. No. 0.019;
Matches   47;  Conservative  28;  Mismatches  60;  Indels  45;  Gaps  9;

Qy    35 TPEPQPTRVILFNKPYDVL--PQFTDEAGR----STL---------KDFIPVQGVYAA  77
         ||     || ||| | |     :||||||    ::!!          || ||||:||
Db    79 TPVPMKLDVLYEDDHYLVLNKPNIAIHPGKGVHVATLIEGLLYYGQEKGFSP----FLV 134

Qy    78 GRLRDRDSEGLLVLTNDGVLQARLTQ--PGKRTGKIYYVQVEGEPPDASLAKLRNGVTLND 135
         ||||||::||||:|||| | || :    ||||:|||:|||:|||||:||||||||||||
Db   135 HRLDKETSGLLVVAKNREAARILTELFKGRNIEKEYVLVRGVPEN----------NM 182

Qy   136 GPTLPA-GIERVNEPEWLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTL 194
         | | :  |||: |||| | || :  |   :|:|  ||:||||:|||||||| :||  :
Db   183 KITIPLDGQEAVSIVSV----KPLKD-----VSLLRVKIHTGRKHQIRRHLSQIGFPVV 233
```

```
RESULT  9
Q9X2Y5
ID   Q9X2Y5       PRELIMINARY;      PRT;   1464 AA.
AC   Q9X2Y5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   L302.8 PROTEIN.
GN   L302.8.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MURPHY L., QUAIL M., LAWSON D., HARRIS D., RAJANDREAM M., IVENS A.,
RA   BARRELL B.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031909; CAB42350.1; -.
SQ   SEQUENCE 1464 AA;  155208 MW;  3DEFB783 CRC32;

Query Match           9.2%;  Score 107.5;  DB 5;  Length 1464;
Best Local Similarity  23.5%;  Pred. No. 0.43;
Matches   48;  Conservative  29;  Mismatches  72;  Indels  55;  Gaps  6;

Qy    54 PQFTDEAGRSTLKDFIPVQGVYAAG---RLDRDSEGLLVLTNDGVLQARLTQPGKRTGK 109
         : |: ||          ||||||       |! | ::|      ::  ||: || : ||
Db  1234 PLFSVSPFLSPMPSAVVCAPVYAAGATASSISRRS--ALPLSHGRQLQS---------- 1281

Qy   110 IYYVQVEGEPPDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLW----------PRNPPIR 160
         :    : ||       :||||| :||       :||   :  ||             |  |
Db  1281 ------PGQPSADALAASRAGSTLSWSSSVPVSLSSLEHVHQQYATLHAALQDQRNAITR 1334

Qy   161 ER-----KSIPTSWLKITLYEGRNRQVRRMTAHVGFP 197
         |:       | : :|| :|| ::|:  |!::  !    ||!
Db  1335 YRVLEYFASAGVAYYQVELCSGRRHQIRQHFAQLGFPLVGDARYHAGVVAGHAGTTLGMR 1394

Qy   198 RYAMGSYTLDSLANGEWRDVTPKE 221
         | ::|::|!! || :: |
Db  1395 RAALHAYTVDILATTDGDDLASSE 1418


RESULT 10
Q50765
ID   Q50765       PRELIMINARY;      PRT;    217 AA.
AC   Q50765; 008191;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   LEXA GENE.
GN   LEXA.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   MOVAHED2ADEH F.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   GENTLES S., CHURCHER C.M.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   BARRELL B.G., RAJANDREAM M.A., PARKHILL J.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
```

RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; X91407; CAA62750.1; -.
DR   EMBL; Z96072; CAB09461.1; -.
DR   HSSP; P03033; 1QAA.
DR   PFAM; PF00717; Peptidase_S24; 1.
DR   PRINTS; PR00726; LEXASERPTASE.
SQ   SEQUENCE   217 AA; 23095 MW; 0015670E CRC32;

Query Match   8.2%; Score 95.5; DB 2; Length 217;
Best Local Similarity 24.2%; Pred. No. 0.39;
Matches 40; Conservative 21; Mismatches 73; Indels 31; Gaps 6;

Qy   13 MTKTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKD--FIP   70
            :| || ||:: | |::   :| :| || | |:| :| | :|:
Db   41 LTSTSSVAHQLRTLERKGYLRRDPN-RPRAVNVRGADDAALPPVTEVAGSDALPEPTFVP   99

Qy   71 VQGVYAAG----------RLRDSEG-----LLVLTNDGVLQARLTQPGKRTGKIYY   112
            |:| | ||| ||:   :||:|
Db   100 VLGRIAAGGPILAEEAVEDVFPLPRELVGEGTLFLLKVIGDSMVEAAICD-----GDWVV 154

Qy   113 VQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNP   157
            :| : |: :| || || ||
Db   155 VRQNVADNGDIV----AAMIDGEATVKTFKRAGGQVWLMPHNP   194

RESULT 11
054664
ID   054664   PRELIMINARY;   PRT;   454 AA.
AC   054664;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE   ORF H1655.
OS   Halobacterium sp.
OG   Plasmid pNRC100.
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 88201675.
RA   JONES J.G., HACKETT N.R., HALLADAY J.T., SCOTHORN D.J., YANG C.F.,
RA   NG W.L., DASSARMA S.;
RA   DASSARMA S., DAMERVAL T., JONES J.G., TANDEAU DE MARSAC N.;
RT   "A plasmid-encoded gas vesicle protein gene in a halophilic
RT   archaebacterium.";
RL   Mol. Microbiol. 1:365-370(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 90016863.
RA   JONES J.G., YOUNG D.C., DASSARMA S.;
RT   "Analysis of insertion mutants reveals two new genes in the pNRC100
RT   gas vesicle gene cluster of Halobacterium halobium.";
RL   Nucleic Acids Res. 17:7785-7793(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 91267967.
RA   NG W.L., KOTHAKOTA S., DASSARMA S.;
RT   "Structure of the gas vesicle plasmid in Halobacterium halobium
RT   inversion isomers, inverted repeats, and insertion sequences.";
RL   J. Bacteriol. 173:3933-3933(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 91323716.

RA   JONES J.G., YOUNG D.C., DASSARMA S.;
RT   "Structure and organization of the gas vesicle gene cluster on the
RT   Halobacterium halobium plasmid pNRC100.";
RL   Gene 102:117-122(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 93012964.
RA   HALLADAY J.T., NG W.L., DASSARMA S.;
RT   "Genetic transformation of a halophilic archaebacterium with a gas
RT   vesicle gene cluster restores its ability to float.";
RL   Gene 119:131-136(1992).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 93139036.
RA   HALLADAY J.T., JONES J.G., LIN F., MACDONALD A.B., DASSARMA S.;
RT   "The rightward gas vesicle operon in Halobacterium plasmid pNRC100:
RT   identification of the gvpA and gvpC gene products by use of antibody
RT   probes and genetic analysis of the region downstream of gvpC.";
RL   J. Bacteriol. 175:684-692(1993).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 93327890.
RA   DASSARMA S.;
RT   "Identification and analysis of the gas vesicle gene cluster on an
RT   unstable plasmid of Halobacterium halobium.";
RL   Experientia 49:482-486(1993).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 93328662.
RA   NG W.L., DASSARMA S.;
RT   "Minimal replication origin of the 200-kilobase Halobacterium plasmid
RT   pNRC100.";
RL   J. Bacteriol. 175:4584-4596(1993).
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE; 95095934.
RA   DASSARMA S., ARORA P., LIN F., MOLINARI E., YIN L.R.;
RT   "Wild-type gas vesicle formation requires at least ten genes in the
RT   gvp gene cluster of Halobacterium halobium plasmid pNRC100.";
RL   J. Bacteriol. 176:7646-7652(1994).
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RA   NG W.L., ARORA P., DASSARMA S.;
RL   Syst. Appl. Microbiol. 16:560-568(1994).
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RA   DASSARMA S.;
RL   (In) Robb F.T., Place A.R., Sowers K.R., Schreier H.J., Dassarma S.,
RL   Fleischmann E.M. (eds.);
RL   Archaea:
RL   A laboratory manual - halophiles, pp. 253-255,
RL   Cold Spring Harbor Laboratory Press, New York (1995).
RN   [12]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RA   DASSARMA S., ARORA P.;
RL   FEMS Microbiol. Lett. 153:1-10(1997).
RN   [13]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RA   NG W.L., CIUFO S.A., SMITH T.M., BUMGARNER R.E., LORETZ C., BASKIN D.,
RA   FAUST J., SETO J., SLAGEL J., HOOD L., DASSARMA S.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [14]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;

```
RA   NG W.L., CIUFO S.A., SMITH T.M., BUMGARNER R.E., LORETZ C., BASKIN D.,
RA   FAUST J., SETO J., SLAGEL J., HOOD L., DASSARMA S.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016485; AAC82930.1; -.
DR   EMBL; AF016485; AAC82847.1; -.
KW   Plasmid.
SQ   SEQUENCE   454 AA;   51384 MW;   F83F7D9B CRC32;

Query Match             7.6%;
Best Local Similarity   20.8%;                    DB 1;  Length 454;
Matches    52;  Conservative   24;  Mismatches   93;  Indels   81;  Gaps  11;

Qy   28  SRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKDF--IPVQG-----VYAAGRL  80
Db  105  SPRLVEEIPDPQPPEITRYNRHY----YQCDSCGTETVAAHPDCPDEGQFGVNVIAQSAL 160

Qy   81  DRDS--------EGLLVLTNDGVLQARLTQPGKRTGKIYYVQVEGEPDDASLAKLR     128
Db  161  SRYDHRLPYRKIADRFEQLHELELSGASAWHATERAAARAGRCEYEQIRREIQDADVVHID 220

Qy  129  NGVTLNDGPTLPAGIERVNEPEWLWPRNP------PIRERKS-----------IP      166
Db  221  E-----------TGIKRDGEQAWIWTFKTAQHTLYAVRESRGSDVPAEVLGEDFAGTVIC  269

Qy  167  TSWLKITLYEGRNRQVRRMTAHV----------------GFPTLRLIR--YAMGSYTLDSLAN 211
Db  270  DGW---TAYPAFSSNLQRCWAHILREAEDAAEKQAEGEPIYHALRQVYVALQARLES---  324

Qy  212  GEWRDVTPKE 221
Db  324  ----DPSPRE 329
```

```
RESULT 12
084108
ID   084108     PRELIMINARY;     PRT;   303 AA.
AC   084108;
DT   01-NOV-1998 (TREMBLrel. 08, Created)
DT   01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE   PREDICTED PSEUDOURIDINE SYNTHETASE FAMILY.
GN   YCEC.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RT   "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT   Chlamydia trachomatis.";
RL   Science 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-D/UW-3/CX;
RA   STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA   MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA   DAVIS R.W.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001284; AAC7697.1; -.
DR   PROSITE; PS01129; PSI_RLU; 1.
DR   PFAM; PF00849; YABO; 1.
SQ   SEQUENCE   303 AA;   34623 MW;   74166AFE CRC32;

Query Match             7.6%;
Best Local Similarity   21.8%;                    DB 2;  Length 303;
Matches    49;  Conservative   28;  Mismatches   88;  Indels   60;  Gaps  10;

Qy    9  PENTMTKTSFRKH------RVERFSS------RQATRRTPEPQP------TRVILFNKP  49
```

```
Db   26  PKKTILE-SVRYHGCRVNGRIERFESYKLQPGDRVSLQIIEHSSPQLLWEDEHLCIYNKP  84

Qy   50  YDVLPQFTDEAGRSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQ--PGKRT 107

Db   85  ------------AKQTSEDLAHQLNVHLVHRLDRDTSGCILFAKHAKASSLITQLFKNREI 133

Qy  108  GKIYYVQVEGEPDDAS--------LAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNPPI  159

Db  134  DKRYIALVFGQPRQESGIITTYTAPCYRRIGAVLFGNTDQNSG--KITITKW-------- 184

Qy  160  RERKSIPTSWLKITL-----YEGRNRQVRRMTAHVGFPTLRLIRY                199

Db  184  ----EILTRYPKYTLLLCRPVTGRTHQIRLHMKTIGHPIVGDVDY                224
```

```
RESULT 13
Q9X2Q7
ID   Q9X2Q7     PRELIMINARY;     PRT;   454 AA.
AC   Q9X2Q7;
DT   01-NOV-1999 (TREMBLrel. 12, Created)
DT   01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE   TBP-LIKE FACTOR (FRAGMENT).
GN   TLF.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DANTONEL J.C., WURTZ J.M., POCH O., MORAS D., TORA L.;
RT   "TBP-like factor: another saddle like structure with a new function in
RT   Metazoans?";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ238443; CAB41469.1; -.
FT   NON_TER    454    454
SQ   SEQUENCE   454 AA;   51209 MW;   78DD85E0 CRC32;

Query Match             7.4%;
Best Local Similarity   21.3%;  Pred. No. 6.6;  DB 5;  Length 454;
Matches    53;  Conservative   30;  Mismatches   97;  Indels   69;  Gaps  10;

Qy    5  QLITPENTMKTSFRKHR-------------VERFSSRQATRRTPEPQPTRVILFNKPYDVLPQ  55

Db  192  QMVQPQ--MTRHQMAQHHAQQPHPQIYVPRDMNLAVPLREPSPEPIPVKI--EVPDVPPE  247

Qy   56  FTDEAGRSTLKD----FIPVQGVYA------------AGRLDRDSEGLLVLTNDGVLQARLTQ 102

Db  248  GTSAANEEPMPDDGDIDIQIRNVVCNYTLPLHIDLRKLAMNTHNVTYEREKGVMMKQKRS  307

Qy  103  PG-----KRTGKIYYVQVEGEPDDASLAKLRNGVTLNDGPTLPAGIERVNEPEWLWPRNP  157

Db  308  PGCYIKVYSSGKVYIVGCRSEADCKRAAR----------SIARHVQRV--------MG   347

Qy  158  PIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGSYTLDSLANGE----  214

Db  348  KYKERVSI----------RNYRVNNVLATCRLPFGIKIEEVAAKYPSESTYEPELSVG    395

Qy  214  --WRDVTPK 220

Db  396  LWRSVTPK  404
```

```
RESULT 14
Q92664
ID   Q92664     PRELIMINARY;     PRT;   473 AA.
AC   Q92664;
DT   01-MAY-1999 (TREMBLrel. 10, Created)
DT   01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 52.7 KD PROTEIN.
OS   Zymomonas mobilis.
OC   Bacteria; Proteobacteria; alpha subdivision; Zymomonas group;
```

```
OC    Zymomonas.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=ZM4;
RA    UM H.W., KANG H.S.;
RT    "Sequence analysis of 43A9 fosmid clone of Zymomonas mobilis ZM4.";
RL    Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF102543; AAD19413.1; -.
KW    Hypothetical protein.
SQ    SEQUENCE    473 AA;    52674 MW;    C886885F CRC32;

Query Match             7.3%;    Score 84.5;  DB 2;   Length 473;
Best Local Similarity  19.8%;    Pred. No. 10;
Matches   50; Conservative   27; Mismatches   60; Indels  115; Gaps  13;

Qy     48 KPYDVLPQFTDEAGRSTLKDFIPVQGVYAAGRLDRDSEGLLVLTNDGVLQARLTQP----  104
          :||||:|||:  :       :||:|   :
Db     99 EPYDPWPDFKNAVNNNTLGDW-----------------------------AANLPPPYSGY  130

Qy    104 -GKRTGKIYY----VQVEGEP----------DDASLAKLRNGVTLND---------G     136
             -|:  :|||    : |   :           :|| ||  : |  |
Db    131 AGLKNGLIIYYRAIAEKGGWPVVPTGADLTVGSDNPRVKTLRRRIAIEDKELPDNGSSHFD  190

Qy    137 PTLPAGIER-----------------------VNEP--------EWLW-PRNPPIRER     162
          |: |:                          ::|:           :| |  |||||
Db    191 AALKAAVKRAQQRYGLNPTGVVSGQTQALNQPVKSRIHQIMANMERWRWLPRNLP-ENR   249

Qy    163 KSIPTSWLKITLYEGRNRQVRRMTAHVG----FPTLRLIRYAMGSYTLDSLANGEW---  215
             :  ||||:|: |:  :||        ::   : |  |  |
Db    250 IQVNVAAAVLTLFE-KNQPVMSMKSVTGRPGNETPMLRSTIYSL------VINPPWNVP  301

Qy    215 -----RDVTPKE  221
               :|| |||
Db    302 DSIAKKELWPKE  313


RESULT  15
Q24717
ID    Q24717         PRELIMINARY;      PRT;     660 AA.
AC    Q24717;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT    01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE    SODIUM CHANNEL PROTEIN (FRAGMENT).
GN    PARA.
OS    Drosophila virilis (Fruit fly).
OC    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC    Ephydroidea; Drosophilidae; Drosophila.
RN    [1]
RP    SEQUENCE FROM N.A.
RA    THACKERAY J.R., GANETZKY B.;
RL    Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 94238327.
RA    THACKERAY J.R., GANETZKY B.;
RT    "Developmentally regulated alternative splicing generates a complex
RT    array of Drosophila para sodium channel isoforms.";
RL    J. Neurosci. 14:2569-2578(1994).
DR    EMBL; U26343; AAB60245.1; -.
DR    FLYBASE; FBgn0015214; Dvir\para.
DR    PFAM; PF00520; ion_trans; 1.
KW    Ionic channel.
FT    NON_TER        1        1
FT    NON_TER      660      660
SQ    SEQUENCE    660 AA;    74460 MW;    DF51E31F CRC32;

Query Match             7.3%;    Score 84.5;  DB 5;   Length 660;
Best Local Similarity  29.9%;    Pred. No. 16;
Matches   50; Conservative   19; Mismatches   19; Indels   25; Gaps   6;
```

```
Qy     11 NTMTKTSFRKHRVERFSSRQATRRTPEPQPTRVILFNKPYDVLPQFTDEAGRSTLKD--F   68
          :||||| | ::|: || | :  ::  : |:: .: |||||:
Db    122 STMTKESKLRNRNTRNQSVGATNGGTTCLDTNHKVEHRDYEIGLECTDEAGKIKHHDNPF  181

Qy     69 I-PVQGVYAAGRLDRDSEGLLVLTNDGVLQA--RLTQPGKRTGKIYYVQVGEPDDASLA  125
          | ||| :  :|| : |:|| | :|
Db    182 IEPVQTQTVV-----DMKDVMVL-NDIIEQAAGRHSRASDRGVSVYYFPTEDDDE-----  231

Qy    126 KLRNGVTLNDGPTL  139
          |||||
Db    231 ---------DGPTL  235


Search completed: June 10, 2000, 11:25:04
Job time: 6515 sec
```

us-09-252-691-7056.rspt

W53871
ID   W53871 standard; peptide; 22 AA.
AC   W53871;
DT   13-JUL-1998 (first entry)
DE   Gravin polypeptide fragment, residues 295-316.
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor.
OS   Homo sapiens.
PN   US5741090-A.
PD   21-APR-1998.
PF   19-DEC-1996; 769309.
PR   19-DEC-1996; US-769309.
PA   (UYOR-) UNIV OREGON HEALTH SCI.
PI   Klauck TM, Nauert JB, Scott JD;
DR   WPI; 98-260552/23.
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
PS   Example 3; Column 10; 32pp; English.
CC   This sequence corresponds to residues 295-316 of gravin, and represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
SQ   Sequence 22 AA;

Query Match   2.7%; Score 6; DB 1; Length 22;
Best Local Similarity   100.0%;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   15 KTSFRK 20
              ||||||
Db   15 KTSFRK 20

RESULT 10
W10941
ID   W10941 standard; peptide; 35 AA.
AC   W10941;
DT   10-NOV-1997 (first entry)
DE   Polyclonal anti-ferritin binder sequence, C28, from R26 library.
KW   Functional surrogate; analyte; affinity receptor; immunoreactive group;
KW   mimic; homogenous immunoassay; detection; diagnostic analyte; Chlamydia;
KW   cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;
KW   pregnancy; infectious disease; ferritin; myosin light chain; troponin;
KW   follicle stimulating hormone; human; growth hormone; immunoglobulin E;
KW   prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;
KW   antibody; chorionic gonadotropin; luteinising hormone; cytomegalovirus;
KW   Streptococcus; rubella; toxoplasma; DK-MB; prostate-specific antigen;
KW   carcinoembryonic antigen; alpha-fetoprotein; herpes virus; CA125.
OS   Synthetic.
PN   WO9641172-A1.
PD   19-DEC-1996.
PF   07-JUN-1996; U10498.
PR   07-JUN-1995; US-476375.
PA   (CYTO-) CYTOGEN CORP.
PI   Carter JM, Lee-Own FV;
DR   WPI; 97-077284/07.
DR   N-PSDB; T48787.
PT   Labelled functional surrogate of an analyte - useful as competitor
PT   molecule in affinity assays, esp. for detecting large macromolecules
PT   such as ferritin
PS   Disclosure; Page 118/2; 156pp; English.
CC   This sequence represents a polyclonal anti-ferritin binder sequence from
CC   the R26 library (C series) which may be used in the conjugate of the

CC   invention. The novel labelled conjugate comprises at least one label
CC   attached to a functional surrogate of an analyte of interest. The
CC   surrogate is capable of competing effectively with the analyte for a
CC   limiting amount of an affinity receptor for the analyte. The conjugate
CC   exhibits an activity that is altered upon interaction with the affinity
CC   receptor and this activity can be measured and related to the amount of
CC   the analyte present in a sample. Functional surrogates such as this have
CC   an immunoreactive group that allows the surrogate to compete effectively
CC   and with the analyte for a limiting amount of its affinity receptor.
CC   Functional surrogates are able to mimic naturally occurring analytes.
CC   They can be labelled for use in standard competitive affinity assays
CC   (esp. homogenous immunoassays) for detecting large macromolecules such
CC   as polypeptides, polysaccharides, polynucleotides, glycoproteins and
CC   lipid-containing macromolecules, as well as small haptens. Typical
CC   diagnostic analytes for detection include cardiac or tumour markers,
CC   allergens, hormones related to fertility-pregnancy or analytes associated
CC   with infectious disease. In particular, the assays are useful for
CC   detecting ferritin, follicle stimulating hormone, human growth hormone,
CC   immunoglobulin E, prolactin, parathyroid hormone, human placental
CC   lactogen, hepatitis antigens or antibodies against them, human
CC   chorionic gonadotropin, human luteinising hormone, cytomegalovirus,
CC   Chlamydia, Streptococcus a, rubella, toxoplasma, herpes virus, DK-MB,
CC   myoglobin, myosin light chain, troponin, carcinoembryonic antigen,
CC   alpha-fetoprotein, prostate-specific antigen and CA125 (a tumour marker).
SQ   Sequence 35 AA;

Query Match   2.7%; Score 6; DB 1; Length 35;
Best Local Similarity   100.0%; Pred.No.35;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   76 AAGRLD 81
              ||||||
Db   17 AAGRLD 22

RESULT 11
W59222
ID   W59222 standard; peptide; 39 AA.
AC   W59222;
DT   27-AUG-1998 (first entry)
DE   Mouse gamma interferon peptide fragment muIFN-gamma (95-133)S.
KW   Interferon-gamma; IFN-gamma; mouse; agonist; intracellular activity;
KW   inducer; macrophage; antiviral; major histocompatability complex;
KW   MHC class II; treatment; disease.
OS   Mus sp.
PN   US5770191-A.
PD   23-JUN-1998.
PF   24-MAY-1995; 449639.
PR   24-MAY-1995; US-449639.
PA   (UYFL) UNIV FLORIDA.
PI   Johnson HM, Szente BE;
DR   WPI; 98-376742/32.
PT   Interferon-gamma agonists with antiviral activity - comprising
PT   C-terminal fragment of interferon-gamma
PS   Example 1; Column 3-4; 13pp; English.
CC   This sequence is a peptide fragment of mouse gamma interferon (IFN-gamma)
CC   which corresponds to amino acids 95 to 133 of the full length protein and
CC   is a scrambled version of the peptide represented in W59221. This peptide
CC   is used in a method which identifies fragments which, once internalised
CC   into a cell act as agonists and have biological activity which is
CC   comparable to the full length IFN-gamma protein. IFN-gamma agonists can
CC   bind to the cytoplasmic domain of the IFN-gamma receptor and exhibit
CC   intracellular interferon-gamma activity, inducing macrophage antiviral
CC   activity and inducing major histocompatability complex (MHC) class II
CC   expression. They can be used to treat a variety of clinically relevant
CC   disease states in animals and humans.
SQ   Sequence 39 AA;

Query Match   2.7%; Score 6; DB 1; Length 39;
Best Local Similarity   100.0%; Pred.No.39;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

CC   bacterial gene expression. The present protein may be used to produce
CC   vaccines to enable a host to produce specific antibodies with
CC   antibacterial action. These vaccines and antibodies would protect
CC   a host against invasion by S. aureus, and conditions relating to
CC   Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC   skin syndrome, and toxic shock syndrome.
SQ   Sequence   71 AA;

Query Match   2.7%;   Score 6;   DB 1;   Length 71;
Best Local Similarity 100.0%;   Pred. No. 69;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   126 KLRNGV 131
             ||||||
Db   36 KLRNGV 41

RESULT 14
W59026
ID   W59026 standard; Protein; 88 AA.
AC   W59026;
DT   14-SEP-1998   (first entry)
DE   Colicin V.
KW   Colicin V; bacteriocin; signal peptide; protein secretion;
KW   preservative; food spoilage; lactic acid bacterium; antibacterial.
OS   Escherichia coli.
FH   Key   Location/Qualifiers
FT   Disulfide_bond   76..87
PN   WO9809639-A1.
PD   12-MAR-1998.
PF   05-SEP-1997; U15609.
PR   05-SEP-1996; US-026257.
PA   (UYAL-) UNIV ALBERTA.
PI   Franz CM, Greer GG, Leisner JJ, McCormick JK, McMullen LM,
PI   Pkns J, Poona, Stiles ME, Van Belkum MJ, Vederas JC,
PI   Worobo RJ, Worobo RW;
DR   WPI; 98-193319/17.
PT   Bacterial growth inhibiting peptide(s) enterocin 900 or brochocin C
PT   - used to inhibit growth of susceptible bacteria in e.g. foodstuff,
PT   living animal, food preparation area or fermentation vessel
PS   Example 3; Page 140; 174pp; English.
CC   This polypeptide comprises colicin V, a bacteriocin produced by
CC   Escherichia coli. Colicin V secretion in Lactococcus lactis hosts
CC   was obtained by fusion to the leucocin A leader peptide (see
CC   W59013). The invention provides secretion vectors in which the
CC   structural and immunity genes of bacteriocins, such as
CC   carnobacteriocin B2 (see W59029-31), colicin V, leucocin A and
CC   brochocin-C (see W59021), are fused to a signal peptide sequence,
CC   such as the colicin V signal peptide, allowing production and
CC   secretion of the bacteriocin(s) by host and heterologous bacteria.
CC   Bacteriocin cassettes encoding two or more bacteriocins are used to
CC   tailor the antibacterial spectrum of a producer strain to target a
CC   range of spoilage or pathogenic bacteria, e.g. E. coli and
CC   Salmonella. Using the technology of multiple bacteriocin
CC   production and delivery using lactic acid bacteria, a range of
CC   bacteriocins can be produced in situ in human food, animal feed,
CC   the gastrointestinal tract of humans and animals, and in the female
CC   genital tract.
SQ   Sequence   88 AA;

Query Match   2.7%;   Score 6;   DB 1;   Length 88;
Best Local Similarity 100.0%;   Pred. No. 84;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   .75 YAAGRL 80
             ||||||
Db   70 YAAGRL 75

RESULT 15
W94884

Qy   20 KHRVER 25
             ||||||
Db   24 KHRVER 29

RESULT 12
R81962
ID   R81962 standard; Protein; 60 AA.
AC   R81962;
DT   18-MAR-1996   (first entry)
DE   ORF encoded by exon 9 of rice pyruvate orthophosphate dikinase gene.
KW   rice; promoter; exon; tissue-specific; season-specific; PPDK;
KW   pyruvate orthophosphate dikinase; C4 photosynthesis; C3 plant.
OS   Oryza sativa L.
PN   J07184657-A.
PD   25-JUL-1995.
PF   28-DEC-1993; 335671.
PF   28-DEC-1993; JP-335671.
PA   (NITA-) JAPAN TAFU GURASU KK.
DR   WPI; 95-287968/38.
DR   N-PSDB; Q98249.
PT   Pyruvate orthophosphate dikinase (C4 photosynthesis-related) gene
PT   and promoter in rice - shows tissue-specific and seasonal specific
PT   expression.
PS   Claim 2; Page 9; 26pp; Japanese.
CC   The sequence is the open reading frame encoded by exon 9 of rice pyruvate
CC   orthophosphate dikinase (PPDK) gene, and which encodes precursor peptide.
CC   The gene shows tissue-specific (in rice leaf and rice arista) and
CC   seasonal specific expression. The PPDK gene may give a C4 photosynthesis
CC   function to a C3 plant.
SQ   Sequence   60 AA;

Query Match   2.7%;   Score 6;   DB 1;   Length 60;
Best Local Similarity 100.0%;   Pred. No. 59;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   104 GKRTGK 109
             ||||||
Db   19 GKRTGK 24

RESULT 13
W27822
ID   W27822 standard; Protein; 71 AA.
AC   W27822;
DT   21-JUL-1998   (first entry)
DE   Staphylococcus aureus protein of unknown function.
KW   Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW   Staphylococcal gene; regulatory element; bacterial gene expression;
KW   vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome.
OS   Staphylococcus aureus.
PN   WO9730070-A1.
PD   21-AUG-1997.
PF   19-FEB-1997; U02318.
PF   20-FEB-1996; US-011888.
PA   (SMIK) SMITHKLINE BEECHAM CORP.
PI   Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR   WPI; 97-424969/39.
DR   N-PSDB; T83787.
PT   Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT   to isolate antimicrobial compounds, and in vaccines against S.
PT   aureus infection
PS   Claim 6; Page 294; 989pp; English.
CC   The present sequence represents a Staphylococcus aureus protein of
CC   unknown function. The DNA sequence was isolated from a library of
CC   clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
CC   be used in the construction of ribozymes and antisense sequences to
CC   control the expression of Staphylococcal genes. The DNA sequence is
CC   also useful as a source of regulatory elements for the control of

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 10:46:30 ; Search time 15.54 Seconds
                                          (without alignments)
                                          338.373 Million cell updates/sec

Title: US-09-252-691-7056
Perfect score: 222
Sequence: 1 AIMRQLITPENTMTKTSFRK........SYTLDSLANGEWRDVTPKEN 222

Scoring table: OLIGO
               Gapop 60.0 , Gapext 60.0

Searched: 188963 seqs, 23366106 residues

Word size :   0

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database :   A_Geneseq_36:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                   SUMMARIES

Result No.	Score	% Score Match	Query Length	DB	ID	Description
1	7	3.2	192	1	W55509	H. pylori ORF 04ge
2	7	3.2	208	1	W55364	H. pylori ORF 04ge
3	7	3.2	241	1	W87996	A human MCG18 prot
4	7	3.2	339	1	W71482	Helicobacter polyp
5	7	3.2	352	1	W22307	Lipase modulator f
6	7	3.2	398	1	W55695	H. pylori ORF 13ae
7	7	3.2	416	1	R44434	Subtilisin-like se
8	6	2.7	764	1	R84345	Melon aconitase en
9	6	2.7	22	1	W53871	Gravin polypeptide
10	6	2.7	35	1	W10941	Polyclonal anti-fe
11	6	2.7	39	1	W59222	Mouse gamma interf
12	6	2.7	60	1	R81962	ORF encoded by exo
13	6	2.7	71	1	W27822	Staphylococcus aur
14	6	2.7	88	1	W59026	Colicin V. Bacteri
15	6	2.7	88	1	W94884	Colicin V cleaved
16	6	2.7	94	1	R74035	Synthetic D32.39 e
17	6	2.7	103	1	W59025	Colicin V prepepti
18	6	2.7	103	1	W94883	Colicin v prepepti
19	6	2.7	127	1	W47279	Papillomavirus maj
20	6	2.7	128	1	W47280	Papillomavirus maj
21	6	2.7	128	1	W47281	Papillomavirus maj
22	6	2.7	132	1	W07861	(DSM 10103) human
23	6	2.7	136	1	W07854	(DSM 10097) human
24	6	2.7	165	1	W07636	P. suis leukotoxin
25	6	2.7	171	1	W31710	Alternate transcri
26	6	2.7	210	1	W31709	x25 gene (exons 1-
27	6	2.7	211	1	W29676	Homo sapiens clone
28	6	2.7	226	1	W21019	H. pylori cell env
29	6	2.7	242	1	W31730	Mycobacterium tube
30	6	2.7	264	1	W77717	3-Isopropylmalate
31	6	2.7	292	1	W53862	Gravin polypeptide
32	6	2.7	321	1	W72133	HSV-2 strain SB5 C
33	6	2.7	321	1	W72183	HSV-2 strain SB5 C
34	6	2.7	360	1	P60517	Sequence of bovine
35	6	2.7	365	1	R38283	NANB hepatitis vir
36	6	2.7	368	1	W98043	Lactobacillus delb
37	6	2.7	376	1	W55644	H. pylori ORF 06cp
38	6	2.7	376	1	W85054	Beta subunit of DN
39	6	2.7	386	1	R98521	Xenopus orphan rec
40	6	2.7	394	1	W49681	Open reading frame
41	6	2.7	405	1	W72031	HSV-2 strain SB5 C
42	6	2.7	409	1	W76772	D. immitis ankyrin
43	6	2.7	409	1	W70604	Ankyrin protein fr
44	6	2.7	416	1	W55708	H. pylori ORF hp7e
45	6	2.7	416	1	W98206	H. pylori GHPO 541
46	6	2.7	432	1	W88236	Mouse prothrombina
47	6	2.7	433	1	W76411	Human betac cytopl
48	6	2.7	439	1	W88235	Human prothrombina
49	6	2.7	466	1	R26505	Adrenalin receptor
50	6	2.7	488	1	R55135	Heat resistant ald
51	6	2.7	495	1	W70270	Dirofilaria immiti
52	6	2.7	500	1	W24582	HPV6 L1 protein. S
53	6	2.7	500	1	W10647	HPV11 mutant L1 pr
54	6	2.7	500	1	W18631	Human HPV6 L1 prot
55	6	2.7	500	1	W15113	Human HPV6 L1 prot
56	6	2.7	500	1	W15115	HPV11 L1 protein c
57	6	2.7	500	1	W15112	Human HPV6 L1 prot
58	6	2.7	500	1	W56791	Human HPV6 L1 prot
59	6	2.7	500	1	W48932	Human papillomavir
60	6	2.7	501	1	W26124	HPV11 mutant L1 pr
61	6	2.7	501	1	W26125	HPV11 mutant L1 pr
62	6	2.7	501	1	W26126	HPV11 mutant L1 pr
63	6	2.7	501	1	W26127	HPV11 mutant L1 pr
64	6	2.7	501	1	W26108	HPV11 mutant L1 pr
65	6	2.7	501	1	W26109	HPV11 mutant L1 pr
66	6	2.7	501	1	W26110	HPV11 mutant L1 pr
67	6	2.7	501	1	W26111	HPV11 mutant L1 pr
68	6	2.7	501	1	W26112	HPV11 mutant L1 pr
69	6	2.7	501	1	W26113	HPV11 mutant L1 pr
70	6	2.7	501	1	W26114	HPV11 mutant L1 pr
71	6	2.7	501	1	W26115	HPV11 mutant L1 pr
72	6	2.7	501	1	W26116	HPV11 mutant L1 pr
73	6	2.7	501	1	W26117	HPV11 mutant L1 pr
74	6	2.7	501	1	W26118	HPV11 mutant L1 pr
75	6	2.7	501	1	W26119	HPV11 mutant L1 pr
76	6	2.7	501	1	W26120	HPV11 mutant L1 pr
77	6	2.7	501	1	W26121	HPV11 mutant L1 pr
78	6	2.7	501	1	W26122	HPV11 mutant L1 pr
79	6	2.7	501	1	W26123	HPV11 mutant L1 pr
80	6	2.7	501	1	W10648	HPV11 mutant L1 pr
81	6	2.7	501	1	W10649	HPV11 mutant L1 pr
82	6	2.7	501	1	W10650	HPV11 mutant L1 pr
83	6	2.7	501	1	W26097	HPV11 mutant L1 pr
84	6	2.7	501	1	W26098	HPV11 mutant L1 pr
85	6	2.7	501	1	W26099	HPV11 mutant L1 pr
86	6	2.7	501	1	W26100	HPV11 mutant L1 pr
87	6	2.7	501	1	W26101	HPV11 mutant L1 pr
88	6	2.7	501	1	W26102	HPV11 mutant L1 pr
89	6	2.7	501	1	W26103	HPV11 mutant L1 pr
90	6	2.7	501	1	W26104	HPV11 mutant L1 pr
91	6	2.7	501	1	W26105	HPV11 mutant L1 pr
92	6	2.7	501	1	W26106	HPV11 mutant L1 pr
93	6	2.7	501	1	W26107	HPV11 mutant L1 pr
94	6	2.7	501	1	W24581	HPV11 mutant L1 pr
95	6	2.7	501	1	W10644	HPV11 mutant L1 pr
96	6	2.7	501	1	W10645	HPV6 mutant L1 pro
97	6	2.7	501	1	W10646	HPV6 mutant L1 pro
98	6	2.7	501	1	W15114	HPV11 mutant L1 pr
99	6	2.7	501	1	W15116	Human HPV11 L1 pro
100	6	2.7	501	1	W15117	HPV11 L1 protein c
101	6	2.7	501	1	W15118	HPV11 L1 protein c
102	6	2.7	501	1	W32080	Human HPV6 L1 prot
103	6	2.7	501	1	W32081	Human HPV6 L1 prot
104	6	2.7	501	1	W32082	Human HPV6 L1 prot
105	6	2.7	501	1	W15109	Human HPV6 L1 prot
106	6	2.7	501	1	W15110	Human HPV6 L1 prot

#			Accession			Description
107	6	2.7	W15111	501	1	Human HPV6 L1 prot
108	6	2.7	W18632	501	1	Human HPV11 L1 pro
109	6	2.7	W05843	507	1	Human papillomavir
110	6	2.7	R83016	532	1	Recombinant bovine
111	6	2.7	W94997	557	1	YAK-1 related seri
112	6	2.7	W38335	572	1	Rat kidney injury
113	6	2.7	W33891	586	1	Flea saliva protei
114	6	2.7	W82365	586	1	Flea saliva protei
115	6	2.7	W86291	597	1	Enterobacter sp. S
116	6	2.7	R86290	600	1	Protaminobacter ru
117	6	2.7	R55799	614	1	Human betaine-GABA
118	6	2.7	R89481	614	1	Human betaine/GABA
119	6	2.7	R86287	629	1	Protaminobacter ru
120	6	2.7	R63226	661	1	Mycoplasma gallise
121	6	2.7	R78646	661	1	Melanoma associate
122	6	2.7	R84854	661	1	MART-1 melanoma an
123	6	2.7	R79910	661	1	MART-1 melanoma an
124	6	2.7	W38164	661	1	M.gallisepticum 66
125	6	2.7	W99657	668	1	Pmel17 encoded by
126	6	2.7	W59008	672	1	Staphylococcus aur
127	6	2.7	W94866	717	1	Leucocin A ATP tra
128	6	2.7	W98740	717	1	ABC transporter (1
129	6	2.7	W81784	734	1	H. Pylori GHPO 961
130	6	2.7	W85603	764	1	P. putida cis/tran
131	6	2.7	R20670	847	1	Hexosaminidase enz
132	6	2.7	R23797	865	1	Lipoxygenase. Prod
133	6	2.7	W99656	865	1	Rice lipoxygenase.
134	6	2.7	R78608	866	1	Staphylococcus aur
135	6	2.7	R92529	878	1	Murine IL-3 recept
136	6	2.7	R20982	878	1	Fas sequence from
137	6	2.7	R76708	897	1	Sequence of beta-c
138	6	2.7	R81953	947	1	Recombinant cold-r
139	6	2.7	R90924	947	1	Rice pyruvate orth
140	6	2.7	R76707	947	1	Corn derived pyruv
141	6	2.7	R90923	953	1	Recombinant cold-r
142	6	2.7	R76711	953	1	F. bidentis pyruva
143	6	2.7	R90927	955	1	Recombinant cold-r
144	6	2.7	W72206	955	1	F. brownii cold re
145	6	2.7	R43671	1144	1	HSV-2 strain SB5 C
146	6	2.7	W72020	1148	1	M.leprae rpoB gene
147	6	2.7	W72098	1180	1	HSV-2 strain SB5 C
148	6	2.7	W93972	1180	1	HSV-2 strain SB5 C
149	6	2.7	R67708	1242	1	Human IRS-1 and IR
150	6	2.7	W59358	1243	1	Insulin receptor s
151	6	2.7	R90583	1244	1	Human retinal dege
152	6	2.7	R15444	1290	1	Phospholipase C-ga
153	6	2.7	W31866	1330	1	Swine herpes virus
154	6	2.7	R91296	1343	1	Mouse metastasis-a
155	6	2.7	W31947	1350	1	Drosophila nitric
156	6	2.7	W31867	1381	1	Rat CASPR/p190, a
157	6	2.7	W76776	1382	1	Human metastasis-a
158	6	2.7	W70608	1384	1	Human CASPR/p190,
159	6	2.7	W53863	1745	1	D. Immitis ankyrin
160	6	2.7	R22210	1745	1	Full length ankyri
161	6	2.7	P20037	1780	1	Human gravin polyp
162	6	2.7	R59923	2206	1	True type 3 poliov
163	6	2.7	R22154	2209	1	Sequence encoded b
164	6	2.7	R40119	2938	1	GAP protein Iral.
165	6	2.7	R66995	3011	1	NANBV Hutch c59 is
166	6	2.7	R79232	3011	1	HCV genomic amino
167	6	2.7	W77397	3011	1	Hepatitis C virus
168	6	2.7	W77398	3011	1	HCV sequence. New
169	6	2.7	W98021	3011	1	Hepatitis C virus
170	6	2.7	W98020	3011	1	Hepatitis C virus-
171	6	2.7	R58777	3011	1	Infectious hepatit
172	6	2.7	W09871	3144	1	Infectious hepatit
173	6	2.7	W36887	3144	1	Protein encoded by
174	6	2.7	W44742	3144	1	Human huntingtin.
175	6	2.7	W52849	3144	1	Previously undescr
176	6	2.7	R44431	3413	1	Human huntingtin p
177	6	2.7	W21037	3567	1	A. mediterranei rI
178	6	2.3		3567		eryA region polype
179	5	2.3		6		Lipolytic enzyme o

#					Accession	Description
180	5	2.3	6	1	W23160	Terminal peptide e
181	5	2.3	6	1	W68444	Hydra peptide anal
182	5	2.3	7	1	R91985	Head Activator-der
183	5	2.3	9	1	R80486	Recombinant bovine
184	5	2.3	9	1	R99969	Clah8 B-cell epito
185	5	2.3	9	1	W49654	Human leucocyte an
186	5	2.3	9	1	W92342	Latent fluorescent
187	5	2.3	10	1	R81152	Melanoma-specific
188	5	2.3	10	1	W32781	Human platelet gly
189	5	2.3	10	1	W48193	Conantokin peptide
190	5	2.3	10	1	W49975	Conantokin peptide
191	5	2.3	10	1	W68435	Hydra peptide anal
192	5	2.3	10	1	W90231	Maltose binding pr
193	5	2.3	11	1	R78785	Rac2 (178-188) pep
194	5	2.3	11	1	R91828	Peptide structural
195	5	2.3	11	1	W68438	Hydra peptide anal
196	5	2.3	12	1	R90334	IL-8 epitope 8-31.
197	5	2.3	12	1	W35423	Thrombopoietin rec
198	5	2.3	12	1	W68440	Hydra peptide anal
199	5	2.3	12	1	Y13238	Natural variant of
200	5	2.3	13	1	W09467	Thrombopoietin rec
201	5	2.3	13	1	W33033	Thrombopoietin rec
202	5	2.3	13	1	W35397	Thrombopoietin rec
203	5	2.3	13	1	W35399	Thrombopoietin rec
204	5	2.3	13	1	W35404	Thrombopoietin rec
205	5	2.3	13	1	W35405	Thrombopoietin rec
206	5	2.3	13	1	W35406	Thrombopoietin rec
207	5	2.3	13	1	W35413	Thrombopoietin rec
208	5	2.3	13	1	W35417	Thrombopoietin rec
209	5	2.3	13	1	W35422	Thrombopoietin rec
210	5	2.3	13	1	W60646	Amino acid sequenc
211	5	2.3	14	1	R08365	Atrial peptide der
212	5	2.3	14	1	R08369	Atrial peptide der
213	5	2.3	14	1	R67030	Fibrinolytic pepti
214	5	2.3	14	1	R81129	Human GMPS tryptic
215	5	2.3	14	1	R93525	Class II +5 GST-YE
216	5	2.3	14	1	R93517	Class II +5 GST-YE
217	5	2.3	14	1	R93522	Class II +5 GST-YE
218	5	2.3	14	1	R93515	Class II +5 GST-YE
219	5	2.3	14	1	R93495	Class II +5 GST-FY
220	5	2.3	14	1	R93483	Class II +5 GST-SR
221	5	2.3	14	1	R93490	Class II +5 GST-SR
222	5	2.3	14	1	R93472	Class II +5 GST-FY
223	5	2.3	14	1	W09466	Class II +5 GST-SR
224	5	2.3	14	1	W09482	Thrombopoietin rec
225	5	2.3	14	1	W09462	Thrombopoietin rec
226	5	2.3	14	1	W33029	Thrombopoietin rec
227	5	2.3	14	1	W33031	Thrombopoietin rec
228	5	2.3	14	1	W33031	Thrombopoietin rec
229	5	2.3	14	1	W35396	Thrombopoietin rec
230	5	2.3	14	1	W35398	Thrombopoietin rec
231	5	2.3	14	1	W35400	Thrombopoietin rec
232	5	2.3	14	1	W35401	Thrombopoietin rec
233	5	2.3	14	1	W35402	Thrombopoietin rec
234	5	2.3	14	1	W35403	Thrombopoietin rec
235	5	2.3	14	1	W35407	Thrombopoietin rec
236	5	2.3	14	1	W35408	Thrombopoietin rec
237	5	2.3	14	1	W35412	Thrombopoietin rec
238	5	2.3	14	1	W36633	Thrombopoietin rec
239	5	2.3	14	1	W36647	Thrombopoietin rec
240	5	2.3	14	1	W66732	Peptide chain of c
241	5	2.3	15	1	R03180	Receptor fragment
242	5	2.3	15	1	R39532	Calmodulin inhibit
243	5	2.3	15	1	R64822	HPV-16 L-protein(3
244	5	2.3	15	1	R94781	Peptide from libra
245	5	2.3	15	1	W38985	Peptide resembling
246	5	2.3	15	1	W43459	Mycobacteria sp. h
247	5	2.3	15	1	W41595	Rat FRAG1 protein
248	5	2.3	15	1	R59379	Human huntingtin.
249	5	2.3	16	1	R83681	Fibronectin fragme
250	5	2.3	16	1	R94782	33 kD fibronectin
251	5	2.3	16	1	R94782	Peptide from libra
252	5	2.3	16	1	W30927	Birch pollen aller

No.	Accession					Description
253	W43460	1	16	2.3	5	Mycobacteria sp. h
254	Y05228	1	16	2.3	5	Human Stat6c antig
255	R49413	1	17	2.3	5	IgG heavy chain po
256	W34872	1	17	2.3	5	Human tau protein
257	P90315	1	18	2.3	5	Alpha-helix-formin
258	R68679	1	18	2.3	5	Consensus B cell e
259	W09448	1	18	2.3	5	Thrombopoietin rec
260	W09497	1	18	2.3	5	Thrombopoietin rec
261	W09456	1	18	2.3	5	Thrombopoietin rec
262	W25849	1	18	2.3	5	HIV-1 env protein
263	W33023	1	18	2.3	5	Thrombopoietin rec
264	W36648	1	18	2.3	5	Thrombopoietin rec
265	W36653	1	18	2.3	5	Thrombopoietin rec
266	W36639	1	18	2.3	5	Thrombopoietin rec
267	W67365	1	18	2.3	5	HIV-1 V3 loop pept
268	W99973	1	18	2.3	5	HIV-1 vaccine synt
269	Y04445	1	18	2.3	5	Peptide CLTB-PRI.
270	Y07833	1	18	2.3	5	Human secreted pro
271	W16887	1	19	2.3	5	Helicobacter pylor
272	W16888	1	19	2.3	5	Helicobacter pylor
273	W09492	1	19	2.3	5	Thrombopoietin rec
274	W09457	1	19	2.3	5	Thrombopoietin rec
275	W09458	1	19	2.3	5	Thrombopoietin rec
276	W09608	1	19	2.3	5	Thrombopoietin rec
277	W33024	1	19	2.3	5	Thrombopoietin rec
278	W33025	1	19	2.3	5	Thrombopoietin rec
279	W36759	1	19	2.3	5	Thrombopoietin rec
280	W36651	1	19	2.3	5	Thrombopoietin rec
281	W36643	1	19	2.3	5	Thrombopoietin rec
282	W71597	1	19	2.3	5	Protein tyrosine p
283	R04980	1	20	2.3	5	Papilloma virus ty
284	P61019	1	20	2.3	5	Alpha-fragment. (am
285	W67535	1	20	2.3	5	IGF-1/IGFBP inhibi
286	W95526	1	21	2.3	5	PAK I peptide frag
287	R29723	1	22	2.3	5	CNP analogue (13).
288	R67808	1	22	2.3	5	BAGE tumor rejecti
289	Y10634	1	22	2.3	5	Peptide antigen SE
290	Y11497	1	22	2.3	5	Human 5' EST secre
291	R58462	1	23	2.3	5	TSAR binding domai
292	Y00900	1	23	2.3	5	GMEB protein domal
293	R92290	1	24	2.3	5	Cycloisomaltooligo
294	W48167	1	24	2.3	5	Conantokin peptide
295	W49950	1	24	2.3	5	Conantokin peptide
296	Y00901	1	24	2.3	5	GMEB protein fragm
297	W02137	1	25	2.3	5	Anti-obesity prote
298	W33916	1	25	2.3	5	D1 dopamine recept
299	R25927	1	26	2.3	5	Peptide for treati
300	R92120	1	26	2.3	5	Streptomyces rosel
301	R52287	1	26	2.3	5	Mouse heavy chain
302	W80364	1	26	2.3	5	N-terminal sequenc
303	W02138	1	27	2.3	5	Anti-obesity prote
304	W39968	1	28	2.3	5	Peptide effecting
305	P81534	1	29	2.3	5	Human insulin acce
306	W00808	1	30	2.3	5	Aspergillus arabin
307	W79386	1	30	2.3	5	DNA encoding Staph
308	Y10862	1	31	2.3	5	Amino acid sequenc
309	W19041	1	32	2.3	5	Truncated T1-SP10E
310	Y11536	1	32	2.3	5	Human 5' EST secre
311	R68670	1	33	2.3	5	p24E/V3 consensus
312	R68709	1	33	2.3	5	P24H/PRI chimera,
313	R68710	1	33	2.3	5	PRI/P24H chimera,
314	R68712	1	33	2.3	5	T5/PRI chimera, CL
315	R68713	1	33	2.3	5	PRI/T5 chimera, CL
316	W25879	1	33	2.3	5	Chimaeric T/B cell
317	W25880	1	33	2.3	5	Chimaeric T/B cell
318	W25882	1	33	2.3	5	Chimaeric T/B cell
319	W25883	1	33	2.3	5	Chimaeric T/B cell
320	W25840	1	33	2.3	5	Chimaeric T/B cell
321	W67396	1	33	2.3	5	HIV-1 peptide epit
322	W67398	1	33	2.3	5	HIV-1 peptide epit
323	W67399	1	33	2.3	5	HIV-1 peptide epit
324	W67395	1	33	2.3	5	HIV-1 peptide epit
325	W67356	1	33	2.3	5	HIV-1 peptide epit
326	W98897	1	33	2.3	5	HIV-1 vaccine synt
327	W98899	1	33	2.3	5	HIV-1 vaccine synt
328	W98900	1	33	2.3	5	HIV-1 vaccine synt
329	W98896	1	33	2.3	5	HIV-1 vaccine synt
330	W99964	1	33	2.3	5	HIV-1 vaccine synt
331	R68674	1	34	2.3	5	T1/V3 consensus se
332	R50643	1	34	2.3	5	G-protein coupled
333	W02835	1	34	2.3	5	G-protein coupled
334	W25844	1	34	2.3	5	Chimeric T/B cell
335	W45585	1	34	2.3	5	peptide fragment o
336	W67360	1	34	2.3	5	HIV-1 peptide epit
337	W99968	1	34	2.3	5	HIV-1 vaccine synt
338	R21134	1	35	2.3	5	PND EE228-1. New a
339	R21135	1	35	2.3	5	PND EE228-2. New a
340	R21136	1	35	2.3	5	PND EE228-3. New a
341	R20853	1	35	2.3	5	PND EE322-2. New a
342	R20883	1	35	2.3	5	PND EE378-3. New a
343	R20910	1	35	2.3	5	PND EE533-1. New a
344	R20911	1	35	2.3	5	PND EE533-2. New a
345	R20912	1	35	2.3	5	PND EE533-3. New a
346	R20937	1	35	2.3	5	PND EE660-1. New a
347	R20936	1	35	2.3	5	PND EE639-3. New a
348	R20951	1	35	2.3	5	PND EE669-1. New a
349	R20952	1	35	2.3	5	PND EE669-2. New a
350	R20953	1	35	2.3	5	PND EE669-3. New a
351	R20958	1	35	2.3	5	PND EE6405-3. New
352	R29110	1	35	2.3	5	Group II HIV gp120
353	R29113	1	35	2.3	5	Group II HIV gp120
354	R29117	1	35	2.3	5	Group II HIV gp120
355	R77720	1	35	2.3	5	p24E/V3 consensus
356	R77760	1	35	2.3	5	P24H/PRI chimera 1
357	R77762	1	35	2.3	5	PRI/P24H chimera l
358	R77762	1	35	2.3	5	T5/PRI chimera lys
359	R77763	1	35	2.3	5	PRI/T5 chimera lys
360	W06251	1	35	2.3	5	A. niger endo-beta
361	R77724	1	36	2.3	5	p24E/V3 consensus
362	P81533	1	37	2.3	5	Human insulin acce
363	P81778	1	37	2.3	5	Sequence encoded b
364	Y01183	1	38	2.3	5	Polypeptide fragme
365	P60641	1	39	2.3	5	Rabbit polysacchar
366	W88639	1	39	2.3	5	Secreted protein e
367	W92752	1	40	2.3	5	WO9846729 Seq ID 1
368	Y13049	1	42	2.3	5	Human secreted pro
369	W02152	1	43	2.3	5	BAGE tumour reject
370	W14929	1	43	2.3	5	Homo sapiens liver
371	W00608	1	44	2.3	5	Large insulin frag
372	W21642	1	45	2.3	5	Grapevine leafroll
373	W00078	1	46	2.3	5	Specific Ob protei
374	W00609	1	46	2.3	5	Large monomer frag
375	Y12322	1	46	2.3	5	Human 5' EST secre
376	P82386	1	48	2.3	5	Human splenin (hSP
377	R15041	1	48	2.3	5	Human thymopoietin
378	W02139	1	48	2.3	5	Anti-obesity prote
379	R33866	1	49	2.3	5	Polypeptide pl192
380	W06523	1	49	2.3	5	Lanthionine modifi
381	W27171	1	49	2.3	5	Human recombinant
382	W34398	1	49	2.3	5	Human Met-OB prote
383	R13350	1	50	2.3	5	P1192 HCV antigen
384	R67967	1	50	2.3	5	Human thymopoietin
385	R67968	1	50	2.3	5	Mouse thymopoietin
386	W89330	1	50	2.3	5	Neisseria meningit
387	Y12747	1	51	2.3	5	Human 5' EST secre
388	R20195	1	52	2.3	5	Heat shock protein
389	R94354	1	52	2.3	5	Synthetic human th
390	R93189	1	52	2.3	5	Synthetic thymopoi
391	W00119	1	52	2.3	5	Generic N-terminal
392	W00064	1	52	2.3	5	Generic N-terminal
393	W14555	1	53	2.3	5	Streptococcus pneu
394	Y01191	1	53	2.3	5	Polypeptide fragme
395	W00116	1	54	2.3	5	Generic N-terminal
396	W00047	1	54	2.3	5	Mammalian Ob prote
397	W00060	1	54	2.3	5	N-terminally trunc
398	W00303	1	54	2.3	5	Human generic ob p

No.	2.3	5	Accession	1	Description
399	2.3	5	W00304	1	Human ob protein s
400	2.3	5	W17704	1	Human obese gene p
401	2.3	5	W87854	1	Lacz alpha peptide
402	2.3	5	R26526	1	Bovine TP (from cl
403	2.3	5	W34876	1	Human tau protein
404	2.3	5	Y02689	1	Human secreted pro
405	2.3	5	R32881	1	Encoded by 600bp f
406	2.3	5	W02140	1	Anti-obesity prote
407	2.3	5	R92783	1	Human intestinal t
408	2.3	5	Y02699	1	Human secreted pro
409	2.3	5	W87788	1	Beta-galactosidase
410	2.3	5	Y12211	1	Human 5' EST secre
411	2.3	5	W00115	1	Generic N-terminal
412	2.3	5	W00059	1	N-terminally trunc
413	2.3	5	R68720	1	p24E/LIP/THAI/NYA
414	2.3	5	W00118	1	Generic N-terminal
415	2.3	5	W00063	1	N-terminally trunc
416	2.3	5	W25890	1	Chimaeric T/B cell
417	2.3	5	W30382	1	Fly derived ttk ge
418	2.3	5	W67406	1	HIV-1 peptide epit
419	2.3	5	W98901	1	HIV-1 vaccine synt
420	2.3	5	R68721	1	T1/LIP/THAI/NYA ch
421	2.3	5	W25891	1	Chimaeric T/B cell
422	2.3	5	W67407	1	HIV-1 peptide epit
423	2.3	5	W98902	1	HIV-1 vaccine synt
424	2.3	5	R77770	1	p24E/LIP/THAI/NYA
425	2.3	5	W27971	1	Amino acid sequenc
426	2.3	5	R77771	1	T1/LIP/THAI/NYA ch
427	2.3	5	W02141	1	Anti-obesity prote
428	2.3	5	W00610	1	Large monomer frag
429	2.3	5	W17702	1	Human obese gene p
430	2.3	5	W09416	1	Human G protein ga
431	2.3	5	W44371	1	Synthetic cell sur
432	2.3	5	Y02983	1	Fragment of human
433	2.3	5	R14540	1	Encoded by.Hepatit
434	2.3	5	R14351	1	HCV protease seque
435	2.3	5	R68541	1	Hepatitis C virus
436	2.3	5	W01692	1	HCV protease clone
437	2.3	5	W46391	1	Amino acid sequenc
438	2.3	5	W97603	1	Amino acid sequenc
439	2.3	5	W33908	1	Peptide analogue o
440	2.3	5	Y13122	1	Human secreted pro
441	2.3	5	Y12140	1	Human 5' EST secre
442	2.3	5	Y01181	1	Polypeptide fragme
443	2.3	5	Y00912	1	cdk2 kinase inhibi
444	2.3	5	Y12313	1	Human 5' EST secre
445	2.3	5	P71265	1	Atrial diuretic pe
446	2.3	5	R99968	1	Cladosporium herba
447	2.3	5	R30165	1	UGT1 Exon 4 produc
448	2.3	5	R26876	1	Human intestinal t
449	2.3	5	W00117	1	Generic N-terminal
450	2.3	5	W00062	1	N-terminally trunc
451	2.3	5	W27631	1	Human intestinal t
452	2.3	5	W94646	1	TNF-R extracellula
453	2.3	5	W00114	1	Generic N-terminal
454	2.3	5	W00058	1	N-terminally trunc
455	2.3	5	Y10854	1	Amino acid sequenc
456	2.3	5	W00611	1	Large monomer frag
457	2.3	5	W27173	1	Human recombinant
458	2.3	5	W34400	1	Human Met-OB prote
459	2.3	5	R81957	1	ORF encoded by exo
460	2.3	5	W02142	1	Anti-obesity prote
461	2.3	5	W06550	1	Human colon specif
462	2.3	5	W30081	1	Mouse persephin. G
463	2.3	5	W30082	1	Rat persephin. GDN
464	2.3	5	W46882	1	Protein sequence e
465	2.3	5	W77718	1	Staphylococcus aur
466	2.3	5	W75861	1	Human secretory pr
467	2.3	5	Y12153	1	Human 5' EST secre
468	2.3	5	Y12166	1	Human 5' EST secre
469	2.3	5	W46757	1	Exon trap L48792.
470	2.3	5	W02143	1	Anti-obesity prote
471	2.3	5	W21563	1	LETO rat cholecyst

No.	5	2.3	Len	1	Accession	Description
472	5	2.3	84	1	W68558	S. peucetius dpsG
473	5	2.3	85	1	W00113	Generic N-terminal
474	5	2.3	85	1	W00057	N-terminally trunc
475	5	2.3	85	1	W00612	Large monomer of m
476	5	2.3	85	1	W20148	Helicobacter pylor
477	5	2.3	86	1	W02144	Anti-obesity prote
478	5	2.3	86	1	Y12336	Human 5' EST secre
479	5	2.3	87	1	W81995	Tobacco trehalase
480	5	2.3	88	1	R03094	Escherichia coli M
481	5	2.3	88	1	P70451	Sequence encoded b
482	5	2.3	88	1	R14298	Seroreactive epito
483	5	2.3	88	1	R14299	Seroreactive epito
484	5	2.3	88	1	R32057	Rabbit serum react
485	5	2.3	90	1	R82805	Ubiquitin conjugat
486	5	2.3	91	1	W38599	S. pneumoniae SARP
487	5	2.3	93	1	P90148	Sequence of hepati
488	5	2.3	93	1	P92031	Sequence encoded i
489	5	2.3	93	1	R14541	Encoded by Hepatit
490	5	2.3	93	1	R14352	HCV protease seque
491	5	2.3	93	1	R68542	Hepatitis C virus
492	5	2.3	93	1	W00112	Generic N-terminal
493	5	2.3	93	1	W00056	N-terminally trunc
494	5	2.3	93	1	W01691	HCV protease clone
495	5	2.3	93	1	W46392	Amino acid sequenc
496	5	2.3	93	1	W97604	Amino acid sequenc
497	5	2.3	94	1	W38546	S. pneumoniae 30S
498	5	2.3	95	1	W67492	lamB signal sequen
499	5	2.3	95	1	Y07825	Human secreted pro
500	5	2.3	96	1	Y00914	Cdk2 kinase inhibi
501	5	2.3	97	1	W00613	Large monomer of m
502	5	2.3	97	1	Y11334	S. pneumoniae ATP-
503	5	2.3	98	1	Y13200	Human secreted pro
504	5	2.3	99	1	W02145	Anti-obesity prote
505	5	2.3	99	1	W10030	Protein encoded by
506	5	2.3	100	1	P80822	Sequence of art pr
507	5	2.3	100	1	W59427	Human chemokine pr
508	5	2.3	101	1	R20607	ART gene protein.
509	5	2.3	101	1	R81437	Hepatitis GB virus
510	5	2.3	102	1	W41157	Biological clock c
511	5	2.3	103	1	W00614	Large monomer of m
512	5	2.3	103	1	W48213	Conus sulcatus con
513	5	2.3	103	1	W49992	Conus sulcatus con
514	5	2.3	104	1	W02146	Anti-obesity prote
515	5	2.3	104	1	W14493	Urease subunit B.
516	5	2.3	104	1	W59497	Mouse C-mu CH1 exo
517	5	2.3	105	1	W00111	Generic N-terminal
518	5	2.3	105	1	W00055	N-terminally trunc
519	5	2.3	105	1	W98746	Biologically activ
520	5	2.3	105	1	R98745	Generic biological
521	5	2.3	105	1	W00615	Large monomer of m
522	5	2.3	105	1	W74873	Human secreted pro
523	5	2.3	106	1	R92521	RecA fragment enco
524	5	2.3	107	1	P95449	Sequence encoded b
525	5	2.3	107	1	R26528	Fetal calf TP (fro
526	5	2.3	107	1	W55337	H. pylori ORF 07ce
527	5	2.3	108	1	W02147	Anti-obesity prote
528	5	2.3	108	1	W86258	Maize OTE partial
529	5	2.3	109	1	R97664	Adenoid-expressed
530	5	2.3	109	1	W10574	Human chemokine al
531	5	2.3	109	1	W07606	Human chemokine al
532	5	2.3	109	1	W59430	Human chemokine pr
533	5	2.3	109	1	W58901	Leukocyte adhesion
534	5	2.3	110	1	R42847	Mutant elastase fr
535	5	2.3	111	1	P50565	Sequence of polype
536	5	2.3	111	1	W00110	Generic N-terminal
537	5	2.3	111	1	W00054	N-terminally trunc
538	5	2.3	111	1	W28276	Amino acid sequenc
539	5	2.3	112	1	W38185	Arabidopsis SCAREC
540	5	2.3	113	1	W00109	Generic N-terminal
541	5	2.3	113	1	W12166	N-terminally trunc
542	5	2.3	114	1	W00053	N-terminally trunc
543	5	2.3	114	1	W02148	Anti-obesity prote
544	5	2.3	114	1	W77549	Staphylococcus aur
				1	W96712	Epithelial nuetrop

#	Accession		Val	Ver		Description
545	Y06831	1	114	2.3	5	Mouse MAb 6A4 muta
546	P90149	1	115	2.3	5	Sequence of hepati
547	P92032	1	115	2.3	5	Sequence encoded 1
548	P50102	1	116	2.3	5	Atrial natriuretic
549	W02149	1	117	2.3	5	Anti-obesity prote
550	W26543	1	117	2.3	5	Trypanosoma cruzi
551	W00616	1	118	2.3	5	Large monomer of m
552	R92732	1	119	2.3	5	Obesity protein C-
553	W00067	1	119	2.3	5	C-terminally trunc
554	W02150	1	119	2.3	5	Anti-obesity prote
555	W02135	1	119	2.3	5	Anti-obesity prote
556	Y11551	1	119	2.3	5	Nickel-binding per
557	Y11895	1	119	2.3	5	Human 5' EST secre
558	W03946	1	120	2.3	5	DNA fragment vk65.
559	Y13025	1	120	2.3	5	Human secreted pro
560	R13161	1	122	2.3	5	Pregnancy-specific
561	R14539	1	122	2.3	5	Hepatitis C Virus
562	R14350	1	122	2.3	5	HCV protease seque
563	R68540	1	122	2.3	5	Hepatitis C virus
564	R89746	1	122	2.3	5	AFT-1 interacting
565	W01690	1	122	2.3	5	HCV protease clone
566	W46390	1	122	2.3	5	Amino acid sequenc
567	W97602	1	122	2.3	5	Amino acid sequenc
568	W00617	1	123	2.3	5	Large monomer of m
569	W14582	1	124	2.3	5	Streptococcus pneu
570	W53996	1	125	2.3	5	Anti-CD4 antibody
571	Y12924	1	125	2.3	5	Amino acid sequenc
572	P60017	1	126	2.3	5	Sequence of cardio
573	R50977	1	126	2.3	5	Norwalk virus RNA-
574	W00108	1	126	2.3	5	Generic N-terminal
575	W00052	1	126	2.3	5	N-terminally trunc
576	Y04862	1	126	2.3	5	Streptococcus pneu
577	Y12656	1	126	2.3	5	Human 5' EST secre
578	R12298	1	127	2.3	5	Norwalk virus RNA-
579	W00618	1	127	2.3	5	Large monomer of m
580	W80674	1	127	2.3	5	S. pneumoniae prot
581	Y12005	1	127	2.3	5	Human 5' EST secre
582	Y04862	1	127	2.3	5	Mycobacterium spec
583	P51238	1	128	2.3	5	Sequence of pro-at
584	W00069	1	128	2.3	5	C-terminally trunc
585	W76346	1	129	2.3	5	Papillomavirus mai
586	W22848	1	129	2.3	5	Mortierella alpina
587	W79301	1	130	2.3	5	A Staphylococcus a
588	W71300	1	130	2.3	5	Obesity protein an
589	W00107	1	131	2.3	5	Generic N-terminal
590	W00051	1	131	2.3	5	N-terminally trunc
591	W14583	1	131	2.3	5	Streptococcus pneu
592	W86277	1	131	2.3	5	Blastx output of h
593	R62446	1	132	2.3	5	81C6 light chain v
594	Y10859	1	132	2.3	5	Amino acid sequenc
595	R33951	1	133	2.3	5	CTMO1 VL. Anti-hum
596	W00619	1	133	2.3	5	Large monomer of m
597	W27117	1	133	2.3	5	Feline interleukin
598	W29751	1	133	2.3	5	Anti-HMFG MAb CTMO
599	Y11226	1	133	2.3	5	S. pneumoniae prot
600	P90154	1	134	2.3	5	Sequence of hepati
601	P92037	1	134	2.3	5	Sequence encoded 1
602	P70564	1	134	2.3	5	Product of ORF 6 f
603	R09420	1	134	2.3	5	B38.1 Light Chain
604	W06209	1	134	2.3	5	MAb B38-1 light ch
605	W14585	1	134	2.3	5	Streptococcus pneu
606	W85056	1	134	2.3	5	Mouse B38.1 light
607	W00106	1	135	2.3	5	Generic N-terminal
608	W00061	1	135	2.3	5	N-terminally trunc
609	W00620	1	136	2.3	5	Large monomer of m
610	W15762	1	136	2.3	5	Cotton fibrous tis
611	W55672	1	136	2.3	5	H. pylori ORF hp6p
612	P81138	1	137	2.3	5	sequence of plant
613	R54978	1	137	2.3	5	Spinach acyl carri
614	W09638	1	137	2.3	5	Oat-derived thionl
615	W28208	1	137	2.3	5	Staphylococcus aur
616	R14542	1	138	2.3	5	Hepatitis C Virus
617	R14353	1	138	2.3	5	HCV protease seque
618	R42211	1	138	2.3	5	CRABP-II gene prod
619	R68543	1	138	2.3	5	Hepatitis C virus
620	W00071	1	138	2.3	5	C-terminally trunc
621	W07434	1	138	2.3	5	Large monomer of m
622	W00606	1	138	2.3	5	Large monomer of m
623	W01689	1	138	2.3	5	HCV protease clone
624	W14774	1	138	2.3	5	HCV protease activ
625	W14775	1	138	2.3	5	HCV protease activ
626	W46393	1	138	2.3	5	Amino acid sequenc
627	W42425	1	138	2.3	5	Pyruvate dehydroge
628	W65025	1	138	2.3	5	Thuja plicata diri
629	W97605	1	138	2.3	5	Amino acid seqeunc
630	Y04922	1	138	2.3	5	Mycobacterium spec
631	W25686	1	139	2.3	5	E. coli beta-galac
632	W70271	1	139	2.3	5	Brugia malayi tran
633	W76342	1	139	2.3	5	Papillomavirus mai
634	W98012	1	139	2.3	5	Hypersensitive res
635	Y04886	1	139	2.3	5	Mycobacterium spec
636	Y04891	1	140	2.3	5	Mycobacterium spec
637	R30454	1	141	2.3	5	C242:11 MAb kappa
638	R32541	1	141	2.3	5	C@42 kappa chain v
639	W00105	1	141	2.3	5	Generic ob protein
640	W00050	1	141	2.3	5	N-terminally trunc
641	R91882	1	141	2.3	5	Keratinocyte growt
642	R94407	1	141	2.3	5	Met-KGF delta-N23/
643	R94404	1	141	2.3	5	Met-KGF delta-N23/
644	R94405	1	141	2.3	5	Met-KGF delta-N23/
645	R94406	1	141	2.3	5	Met-KGF delta-N23/
646	W00353	1	141	2.3	5	KGF delta-N23/R(14
647	R61424	1	141	2.3	5	Keratinocyte growt
648	P61688	1	142	2.3	5	Sequence of enzyme
649	P93370	1	143	2.3	5	protein sequence e
650	P50063	1	143	2.3	5	Soybean heat shock
651	W98594	1	143	2.3	5	H. pylori GHPO 115
652	R06431	1	144	2.3	5	N-terminal of SP1-
653	R47247	1	144	2.3	5	Fragment of 101 kD
654	W00049	1	144	2.3	5	N-terminally trunc
655	W03523	1	144	2.3	5	Generic anti-obesi
656	W03524	1	144	2.3	5	Anti-obesity prote
657	W47082	1	144	2.3	5	Salmonella Sef14 p
658	W98379	1	144	2.3	5	H. pylori GHPO 143
659	W00541	1	145	2.3	5	Human mature obesi
660	W00302	1	145	2.3	5	Human delta Gln28
661	W31902	1	145	2.3	5	Human cystatin F.
662	W32323	1	145	2.3	5	Mature human cysta
663	W30893	1	145	2.3	5	Synthetic obesity
664	W00539	1	146	2.3	5	Human mature obesi
665	W00013	1	146	2.3	5	Acid stable modifi
666	R99490	1	146	2.3	5	Chimeric ob protei
667	R99491	1	146	2.3	5	Chimeric ob protei
668	W00012	1	146	2.3	5	Acid stable modifi
669	R99492	1	146	2.3	5	Chimeric ob protei
670	R99493	1	146	2.3	5	Chimeric ob protei
671	R99494	1	146	2.3	5	Chimeric ob protei
672	R99495	1	146	2.3	5	Chimeric ob protei
673	W00014	1	146	2.3	5	Chimeric ob protei
674	W00015	1	146	2.3	5	Chimeric ob protei
675	W00016	1	146	2.3	5	Chimeric ob protei
676	R99489	1	146	2.3	5	Generic ob protein
677	R99499	1	146	2.3	5	Acid stable modifi
678	W00011	1	146	2.3	5	Acid stable modifi
679	R99500	1	146	2.3	5	Acid stable modifi
680	W00010	1	146	2.3	5	Acid stable modifi
681	W00301	1	146	2.3	5	Human ob protein.
682	W06524	1	146	2.3	5	Wild type ob prote
683	W22900	1	146	2.3	5	Biologically activ
684	W10151	1	146	2.3	5	Biologically activ
685	W22901	1	146	2.3	5	Properly folded ob
686	W10152	1	146	2.3	5	Properly folded ob
687	W22902	1	146	2.3	5	Biologically activ
688	W10153	1	146	2.3	5	Properly folded ob
689	W22903	1	146	2.3	5	Biologically activ
690	W08599	1	146	2.3	5	Properly folded ob

No.					Accession	Description
691	5	2.3	146	1	W08595	Properly folded ob
692	5	2.3	146	1	W22896	Biologically activ
693	5	2.3	146	1	W11279	Generic leptin mim
694	5	2.3	146	1	W11278	Generic leptin mim
695	5	2.3	146	1	W11275	Human leptin for t
696	5	2.3	146	1	W18621	Obesity protein an
697	5	2.3	146	1	W18622	Obesity protein an
698	5	2.3	146	1	W18623	Obesity protein an
699	5	2.3	146	1	W18630	Obesity protein an
700	5	2.3	146	1	W18624	Obesity protein an
701	5	2.3	146	1	W18625	Obesity protein an
702	5	2.3	146	1	W18626	Obesity protein an
703	5	2.3	146	1	W18627	Obesity protein an
704	5	2.3	146	1	W18628	Obesity protein an
705	5	2.3	146	1	W18629	Obesity protein an
706	5	2.3	146	1	W24031	Human obesity prot
707	5	2.3	146	1	W13838	Chimpanzee leptin.
708	5	2.3	146	1	W13839	Gorilla leptin. Re
709	5	2.3	146	1	W13840	Orangutan leptin.
710	5	2.3	146	1	W30727	Haemoglobin beta c
711	5	2.3	146	1	W32568	Anti obesity prote
712	5	2.3	146	1	W32569	Anti obesity prote
713	5	2.3	146	1	W32570	Anti obesity prote
714	5	2.3	146	1	W32571	Anti obesity prote
715	5	2.3	146	1	W32572	Anti obesity prote
716	5	2.3	146	1	W32573	Anti obesity prote
717	5	2.3	146	1	W32574	Anti obesity prote
718	5	2.3	146	1	W32575	Anti obesity prote
719	5	2.3	146	1	W32576	Anti obesity prote
720	5	2.3	146	1	W32577	Anti obesity prote
721	5	2.3	146	1	W32578	Anti obesity prote
722	5	2.3	146	1	W32579	Anti obesity prote
723	5	2.3	146	1	W32580	Anti obesity prote
724	5	2.3	146	1	W26189	Anti obesity prote
725	5	2.3	146	1	W26190	Anti obesity prote
726	5	2.3	146	1	W26191	Obesity protein an
727	5	2.3	146	1	W26192	Obesity protein an
728	5	2.3	146	1	W26193	Obesity protein an
729	5	2.3	146	1	W26194	Obesity protein an
730	5	2.3	146	1	W26195	Obesity protein an
731	5	2.3	146	1	W26196	Obesity protein an
732	5	2.3	146	1	W26197	Obesity protein an
733	5	2.3	146	1	W26198	Obesity protein an
734	5	2.3	146	1	W26199	Obesity protein an
735	5	2.3	146	1	W26200	Obesity protein an
736	5	2.3	146	1	W26201	Obesity protein an
737	5	2.3	146	1	W34483	Human obesity prot
738	5	2.3	146	1	W34489	Obesity protein an
739	5	2.3	146	1	W34490	Obesity protein an
740	5	2.3	146	1	W34491	Obesity protein an
741	5	2.3	146	1	W34492	Obesity protein an
742	5	2.3	146	1	W34493	Obesity protein an
743	5	2.3	146	1	W34494	Obesity protein an
744	5	2.3	146	1	W34495	Obesity protein an
745	5	2.3	146	1	W34496	Obesity protein an
746	5	2.3	146	1	W34482	Human obesity prot
747	5	2.3	146	1	W34484	Obesity protein an
748	5	2.3	146	1	W34485	Obesity protein an
749	5	2.3	146	1	W34486	Obesity protein an
750	5	2.3	146	1	W34487	Obesity protein an
751	5	2.3	146	1	W34488	Obesity protein an
752	5	2.3	146	1	W30791	Obesity protein co
753	5	2.3	146	1	W34396	Human Met-OB prote
754	5	2.3	146	1	W34397	Human Met-OB prote
755	5	2.3	146	1	W30898	Synthetic obesity
756	5	2.3	146	1	W30899	Synthetic obesity
757	5	2.3	146	1	W30900	Synthetic obesity
758	5	2.3	146	1	W36448	Synthetic obesity
759	5	2.3	146	1	W30896	Synthetic obesity
760	5	2.3	146	1	W30897	Synthetic obesity
761	5	2.3	146	1	W30892	Synthetic obesity
762	5	2.3	146	1	W30894	Synthetic obesity
763	5	2.3	146	1	W30895	Synthetic obesity

No.					Accession	Description
764	5	2.3	146	1	W45473	Anti-obesity prote
765	5	2.3	146	1	W45474	Anti-obesity prote
766	5	2.3	146	1	W45476	Anti-obesity prote
767	5	2.3	146	1	W45478	Anti-obesity prote
768	5	2.3	146	1	W53329	Obesity protein an
769	5	2.3	146	1	W53330	Obesity protein an
770	5	2.3	146	1	W53331	Obesity protein an
771	5	2.3	146	1	W53332	Obesity protein an
772	5	2.3	146	1	W53333	Obesity protein an
773	5	2.3	146	1	W53334	Obesity protein an
774	5	2.3	146	1	W53335	Obesity protein an
775	5	2.3	146	1	W53336	Obesity protein an
776	5	2.3	146	1	W53337	Obesity protein an
777	5	2.3	146	1	W53338	Obesity protein an
778	5	2.3	146	1	W53339	Obesity protein an
779	5	2.3	146	1	W53340	Obesity protein an
780	5	2.3	146	1	W53341	Obesity protein an
781	5	2.3	146	1	W53342	Obesity protein an
782	5	2.3	146	1	W50999	Anti-obesity prote
783	5	2.3	146	1	W51000	Anti-obesity prote
784	5	2.3	146	1	W51001	Anti-obesity prote
785	5	2.3	146	1	W51002	Anti-obesity prote
786	5	2.3	146	1	W51003	Anti-obesity prote
787	5	2.3	146	1	W51004	Anti-obesity prote
788	5	2.3	146	1	W51005	Anti-obesity prote
789	5	2.3	146	1	W50993	Anti-obesity prote
790	5	2.3	146	1	W50994	Anti-obesity prote
791	5	2.3	146	1	W50995	Anti-obesity prote
792	5	2.3	146	1	W50996	Anti-obesity prote
793	5	2.3	146	1	W50998	Anti-obesity prote
794	5	2.3	146	1	W57258	Anti-obesity prote
795	5	2.3	146	1	W62303	Amino acid sequenc
796	5	2.3	146	1	W62304	Anti-obesity prote
797	5	2.3	146	1	W62305	Anti-obesity prote
798	5	2.3	146	1	W62306	Anti-obesity prote
799	5	2.3	146	1	W62307	Anti-obesity prote
800	5	2.3	146	1	W62308	Anti-obesity prote
801	5	2.3	146	1	W62309	Anti-obesity prote
802	5	2.3	146	1	W62310	Anti-obesity prote
803	5	2.3	146	1	W62601	Anti-obesity prote
804	5	2.3	146	1	W62602	Formula I for an o
805	5	2.3	146	1	W62603	Formula II for an
806	5	2.3	146	1	W62605	Formula III for an
807	5	2.3	146	1	W62606	Preferred obesity
808	5	2.3	146	1	W62607	Preferred obesity
809	5	2.3	146	1	W62608	Preferred obesity
810	5	2.3	146	1	W62609	Preferred obesity
811	5	2.3	146	1	W62610	Preferred obesity
812	5	2.3	146	1	W62604	Preferred obesity
813	5	2.3	146	1	W49019	Formula IV for an
814	5	2.3	146	1	W49020	Mutant human obese
815	5	2.3	146	1	W49021	Mutant human obese
816	5	2.3	146	1	W49022	Mutant human obese
817	5	2.3	146	1	W62611	Preferred obesity
818	5	2.3	146	1	W57816	Anti-obesity prote
819	5	2.3	146	1	W57817	Anti-obesity prote
820	5	2.3	146	1	W57818	Anti-obesity prote
821	5	2.3	146	1	W57819	Anti-obesity prote
822	5	2.3	146	1	W57813	Anti-obesity prote
823	5	2.3	146	1	W57814	Anti-obesity prote
824	5	2.3	146	1	W57815	Anti-obesity prote
825	5	2.3	146	1	W59919	Leptin receptor ag
826	5	2.3	146	1	W70369	Leptin receptor li
827	5	2.3	146	1	W69682	Human obesity prot
828	5	2.3	146	1	W69683	Obesity protein an
829	5	2.3	146	1	W69684	Obesity protein an
830	5	2.3	146	1	W69686	Obesity protein an
831	5	2.3	146	1	W69687	Obesity protein an
832	5	2.3	146	1	W69688	Obesity protein an
833	5	2.3	146	1	W69685	Obesity protein an
834	5	2.3	146	1	W69689	Obesity protein an
835	5	2.3	146	1	W69690	Obesity protein an
836	5	2.3	146	1	W69691	Obesity protein an

ID	Accession		Num			Description
837	W69692	1	146	2.3	5	Obesity protein an
838	W69693	1	146	2.3	5	Obesity protein an
839	W69694	1	146	2.3	5	Obesity protein an
840	W69695	1	146	2.3	5	Obesity protein an
841	W71304	1	146	2.3	5	Obesity protein an
842	W71305	1	146	2.3	5	Obesity protein an
843	W71306	1	146	2.3	5	Obesity protein an
844	W71307	1	146	2.3	5	Obesity protein an
845	W71308	1	146	2.3	5	Obesity protein an
846	W71309	1	146	2.3	5	Obesity protein an
847	W71310	1	146	2.3	5	Obesity protein an
848	W71311	1	146	2.3	5	Obesity protein an
849	W71298	1	146	2.3	5	Obesity protein an
850	W71299	1	146	2.3	5	Obesity protein an
851	W71301	1	146	2.3	5	Obesity protein an
852	W71302	1	146	2.3	5	Obesity protein an
853	W71303	1	146	2.3	5	Obesity protein an
854	W71833	1	146	2.3	5	Obesity protein an
855	W71840	1	146	2.3	5	Obesity protein an
856	W71841	1	146	2.3	5	Obesity protein an
857	W71842	1	146	2.3	5	Obesity protein an
858	W71843	1	146	2.3	5	Obesity protein an
859	W71844	1	146	2.3	5	Obesity protein an
860	W71845	1	146	2.3	5	Obesity protein an
861	W71834	1	146	2.3	5	Obesity protein an
862	W71835	1	146	2.3	5	Obesity protein an
863	W71836	1	146	2.3	5	Obesity protein an
864	W71837	1	146	2.3	5	Obesity protein an
865	W71838	1	146	2.3	5	Obesity protein an
866	W71839	1	146	2.3	5	Obesity protein an
867	W80501	1	146	2.3	5	Leptin receptor an
868	W80499	1	146	2.3	5	Leptin receptor an
869	R05131	1	146	2.3	5	CF7 peptide encode
870	R92735	1	147	2.3	5	Recombinant human
871	W08383	1	147	2.3	5	Recombinant human
872	W27170	1	147	2.3	5	Human recombinant
873	W27167	1	147	2.3	5	Human recombinant
874	W27169	1	147	2.3	5	Human recombinant
875	W27166	1	147	2.3	5	Human recombinant
876	W22435	1	147	2.3	5	Human Met-OB prote
877	W34394	1	147	2.3	5	Human Met-OB prote
878	W27597	1	147	2.3	5	Recombinant human
879	W53328	1	147	2.3	5	Human obesity prot
880	W52832	1	147	2.3	5	Human OB protein.
881	W49072	1	147	2.3	5	Recombinant human
882	Y13074	1	147	2.3	5	Human secreted pro
883	R99496	1	148	2.3	5	Generic Met-X-ob p
884	W07192	1	148	2.3	5	Human anti-obesity
885	W28801	1	148	2.3	5	Human mast protein
886	W62312	1	148	2.3	5	Anti-obesity prote
887	W62313	1	148	2.3	5	Anti-obesity prote
888	W62314	1	148	2.3	5	Anti-obesity prote
889	W62315	1	148	2.3	5	Human ob protein S
890	W62311	1	148	2.3	5	Anti-obesity prote
891	W57820	1	148	2.3	5	Anti-obesity prote
892	W57821	1	148	2.3	5	Anti-obesity prote
893	W57822	1	148	2.3	5	Anti-obesity prote
894	W57823	1	148	2.3	5	Anti-obesity prote
895	W57824	1	148	2.3	5	Anti-obesity prote
896	W57825	1	148	2.3	5	Anti-obesity prote
897	W71848	1	148	2.3	5	Obesity protein an
898	W71849	1	148	2.3	5	Obesity protein an
899	W62901	1	149	2.3	5	Mutant of the firs
900	W38628	1	149	2.3	5	Streptococcus pneu
901	R11976	1	150	2.3	5	Ferredoxin FdI clo
902	R11977	1	150	2.3	5	Ferredoxin FdI clo
903	R39387	1	150	2.3	5	Protien encoded by
904	W62790	1	150	2.3	5	Amino acid sequenc
905	P51237	1	151	2.3	5	Sequence of pre-pr
906	P60325	1	152	2.3	5	Ggamma-rat atrial
907	P51242	1	152	2.3	5	Sequence of pre-pr
908	R36936	1	152	2.3	5	Rat pre-proANVp. N
909	R74344	1	152	2.3	5	BRK-1 soluble frag

Description	ID	Accession	Num		
Mouse soluble BMP	910	R95229	152	2.3	5
Rat atrial natriur	911	W98191	152	2.3	5
Deacetylcephalospo	912	R13016	153	2.3	5
Soybean heat shock	913	P50065	153	2.3	5
Wound-inducible po	914	R20099	153	2.3	5
Soybean heat shock	915	P50066	154	2.3	5
Polyhydroxyalkanoa	916	R74762	155	2.3	5
H. pylori derived	917	W20632	155	2.3	5
Partial human non-	918	W37255	155	2.3	5
Papillomavirus mai	919	W76348	155	2.3	5
ORFI gene prod. in	920	R63121	158	2.3	5
Mabinlin MBlIII fr	921	W23588	158	2.3	5
S. pneumoniae poss	922	W38512	158	2.3	5
Partial 36 kD anti	923	R05583	159	2.3	5
Sequence of Mycoba	924	P80214	159	2.3	5
MTB protein. Rapid	925	R13395	159	2.3	5
Mycobacterium tube	926	W63034	159	2.3	5
H. pylori GHPO 762	927	W98241	159	2.3	5
H. pylori ORF hpe	928	Y11077	159	2.3	5
Hepatitis B virus	929	W50238	159	2.3	5
Modified bone morp	930	W04233	161	2.3	5
H. pylori surface	931	W20313	162	2.3	5
Streptococcus pneu	932	W61014	163	2.3	5
Keratinocyte growt	933	R91881	164	2.3	5
Met-KGF R(144)Q. K	934	R94401	164	2.3	5
Met-KGF C(1.15)S/R	935	R94403	164	2.3	5
KGF R(144)Q. Purif	936	W00349	164	2.3	5
KGF C(1,15)S/R(144	937	W00350	164	2.3	5
Keratinocyte growt	938	W61423	164	2.3	5
SefA sequence. Eli	939	R62752	165	2.3	5
Human obesity prot	940	W00505	165	2.3	5
Salmonella enterit	941	W23571	165	2.3	5
Human obesity prot	942	W00525	166	2.3	5
Human obesity prot	943	W00515	166	2.3	5
Human obesity prot	944	W00530	166	2.3	5
Human obesity prot	945	W00531	166	2.3	5
Human obesity prot	946	W00533	166	2.3	5
Human obesity prot	947	W00535	166	2.3	5
Human obesity prot	948	W00523	166	2.3	5
Human obesity prot	949	W00532	166	2.3	5
Human obesity prot	950	R68621	167	2.3	5
Polypeptide with s	951	W00516	167	2.3	5
Human obesity prot	952	W00518	167	2.3	5
Human obesity prot	953	W00521	167	2.3	5
Human obesity prot	954	W00519	167	2.3	5
Human obesity prot	955	W00520	167	2.3	5
Human obesity prot	956	W00517	167	2.3	5
Obesity protein. O	957	R92720	167	2.3	5
Human ob protein.	958	R99473	167	2.3	5
Human obese (ob) p	959	W03694	167	2.3	5
Human obese (ob) p	960	W34060	167	2.3	5
Human leptin seque	961	W57442	167	2.3	5
Mouse IMC carcinom	962	W98910	167	2.3	5
Human adipogenesis	963	R43261	169	2.3	5
Human skeletal mus	964	W37496	170	2.3	5
Human spermidine/s	965	W58394	170	2.3	5
S. pneumoniae SGHR	966	W38722	170	2.3	5
Human HUCE-1 prote	967	W92251	170	2.3	5
HCV NS2-NS4 peptid	968	R29855	171	2.3	5
Amino acid sequenc	969	Y12959	171	2.3	5
Part of human alph	970	R05412	172	2.3	5
HCV NS2-NS4 peptid	971	R29560	174	2.3	5
HCV NS2-NS4 peptid	972	R29854	174	2.3	5
HCV NS2-NS4 peptid	973	R29856	174	2.3	5
Hepatitis C virus	974	W13768	174	2.3	5
Hepatitis C virus	975	W13770	174	2.3	5
Human herpesvirus	976	W40102	175	2.3	5
Calcineurin regula	977	Y00881	175	2.3	5
Fimbrial antigen.	978	R23731	176	2.3	5
SE fimbrial antige	979	R42173	176	2.3	5
Streptococcus pneu	980	W62739	176	2.3	5
Human adipogenesis	981	R43262	177	2.3	5
Hepatitis GB virus	982	R81433	177	2.3	5

983	5	2.3	R14890	1	178	1	Human macrophage c
984	5	2.3	R26181	1	178	1	TSG-14. Tumour nec
985	5	2.3	R75337	1	178	1	Human interleukin-
986	5	2.3	W02202	1	178	1	Human interleukin-
987	5	2.3	W60283	1	178	1	Modified xylanase
988	5	2.3	W60742	1	178	1	Xylanase I of Tric
989	5	2.3	W69734	1	178	1	Human cystatin-lik
990	5	2.3	R26529	1	179	1	Fetal calf TP-like
991	5	2.3	R63284	1	180	1	Polypeptide encode
992	5	2.3	R63277	1	180	1	Polypeptide encode
993	5	2.3	R63278	1	180	1	Polypeptide encode
994	5	2.3	R63279	1	180	1	Polypeptide encode
995	5	2.3	R63280	1	180	1	Polypeptide encode
996	5	2.3	R63281	1	180	1	Polypeptide encode
997	5	2.3	R63282	1	180	1	Polypeptide encode
998	5	2.3	R63283	1	180	1	Polypeptide encode
999	5	2.3	W55490	1	180	1	H. pylori ORF hp5p
1000	5	2.3	W47081	1	180	1	Salmonella Sefl4 p

ALIGNMENTS

RESULT   1
W55509
ID W55509 standard; Protein; 192 AA.
AC W55509;
DT 30-JUN-1998 (first entry)
DE H. pylori ORF 04gel0816_22208653l_f2_10 cytoplasmic protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW bacterium.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB; V24918.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,96; Pages 712-713; 1145pp; English.
CC This sequence is a H. pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The DNA and probes derived from it may be used for the
CC identification of H. pylori in a sample, and the diagnosis of
CC H. pylori infection. Nucleic acid sequences complementary to the
CC DNA act as antisense sequences, and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can
CC be used in immunoassays to evaluate the abundance and distribution
CC of H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were
CC analysed for ORF of at least 180 nucleotides, and the predicted
CC coding regions defined by computer evaluation. To identify likely
CC H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having
CC identified and determined the sequences of interest, particular
CC regions can be isolated from H. pylori by PCR amplification for
CC recombinant polypeptide production, e.g. in E. coli hosts.
SQ Sequence   192 AA;

Query Match
Best Local Similarity   3.2%;   Score 7;   DB 1;   Length 192;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   128 RNGVTLN 134
          |||||||
Db   4 RNGVTLN 10

RESULT   2
W55364
ID W55364 standard; Protein; 208 AA.
AC W55364;
DT 17-JUN-1998 (first entry)
DE H. pylori ORF 04gel1210orf1 protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR N-PSDB: V24773.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claim 14; Page 584-585; 1145pp; English.
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported ORF
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence   208 AA;

Query Match
Best Local Similarity   3.2%;   Score 7;   DB 1;   Length 208;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   128 RNGVTLN 134
          |||||||
Db   20 RNGVTLN 26

RESULT   3
W87996
ID W87996 standard; Protein; 241 AA.
AC W87996;
DT 15-APR-1999 (first entry)

```
DE   A human MCG18 protein.
KW   MCG4 protein; gene regulatory function; heat shock protein;
KW   guanine nucleotide exchange factor protein; MCG7 protein;
KW   heat shock-binding protein; MCG18 protein; zinc finger protein;
KW   cancer.
OS   Homo sapiens.
PN   WO9853061-A1.
PD   26-NOV-1998.
PF   22-MAY-1998; AU0380.
PR   22-JAN-1998; AU-001460.
PR   23-MAY-1997; AU-006972.
PR   23-MAY-1997; AU-006973.
PR   23-MAY-1997; AU-006974.
PR   22-JAN-1998; AU-001458.
PR   22-JAN-1998; AU-001459.
PA   (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
PI   Gartside M, Grimmond S, Hancock J, Hayward N, Silins G;
DR   WPI; 99-070146/06.
DR   N-PSDB; X04554.
PT   New gene-expression regulatory genes, mcg4, mcg7, and mcg18 - encode
PT   a zinc finger protein, a GEF, and a heat shock or heat shock binding
PT   protein, useful to detect and treat cancer
PS   Claim 7; Fig 19; 80pp; English.
CC   The present sequence represents a MCG18 protein. The protein has gene
CC   regulatory functions, and has homology to a zinc finger protein.
CC   The specification also describes MCG4, which is homologous to
CC   guanine nucleotide exchange factor protein, and MCG7, which is
CC   homologous to a heat shock protein or heat shock-binding protein.
CC   Detection of mutations in the MCG genes can be used to identify the
CC   propensity for various types of cancer, and to treat, arrest, or
CC   otherwise ameliorate, the effects of a cancer in an animal or bird.
SQ   Sequence    241 AA;

Query Match           3.2%;  Score 7;  DB 1;  Length 241;
Best Local Similarity  100.0%;  Pred. No. 21;
Matches    7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    152 LWPRNPP 158
          |||||||
Db     12 LWPRNPP 18

RESULT 4
W71482
ID   W71482 standard; Protein; 339 AA.
AC   W71482;
DT   09-NOV-1998 (first entry)
DE   Helicobacter polypeptide GHPO 408.
KW   GHPO 408; Infection; therapy; diagnosis; vaccine; gastritis;
KW   ulcer.
OS   Helicobacter pylori.
PN   WO9821225-A1.
PD   22-MAY-1998.
PF   14-NOV-1997; U21353.
PR   29-JUL-1997; US-902615.
PR   14-NOV-1996; US-749051.
PR   01-APR-1997; US-831309.
PR   01-APR-1997; US-833457.
PR   01-APR-1997; US-834705.
PR   24-JUN-1997; US-881227.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI   Al-Garawl A, Haas R, Kleanthous H, Meyer T, Miller C,
PI   Odenbreit S, Tomb J;
DR   WPI; 98-297855/26.
DR   N-PSDB; V52017.
PT   Helicobacter polynucleotide and polypeptide sequences - useful to
PT   treat or prevent gastrointestinal infection
PS   Claim 1; Page 100-101; 362pp; English.
CC   This claimed Helicobacter pylori polypeptide, designated GHPO 408,
CC   can be used in vaccination methods for preventing or treating
CC   Helicobacter infection.    85 Helicobacter polypeptides (see
CC   W71474-W71558) are claimed, as well as isolated polynucleotides
CC   (see V52009-93) that encode them. The invention also provides:
CC   methods for producing these Helicobacter polypeptides in
CC   recombinant host systems, and related expression cassettes, vectors
CC   and transformed or transfected host cells; live vaccine vectors
CC   that contain the polynucleotides of the invention and which can be
CC   used to prevent or treat Helicobacter infection; therapeutic and/or
CC   prophylactic methods involving administration of polynucleotide
CC   molecules, polypeptides or monospecific antibodies; methods for
CC   detecting the presence of Helicobacter in samples using e.g.
CC   the polypeptides or monospecific antibodies; and methods for
CC   purifying the polypeptides by antibody-based affinity
CC   chromatography.
SQ   Sequence    339 AA;

Query Match           3.2%;  Score 7;  DB 1;  Length 339;
Best Local Similarity  100.0%;  Pred. No. 29;
Matches    7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    128 RNGVTLN 134
          |||||||
Db    151 RNGVTLN 157

RESULT 5
W22307
ID   W22307 standard; Protein; 352 AA.
AC   W22307;
DT   18-MAR-1998 (first entry)
DE   Lipase modulator from P. wisconsinensis.
KW   Pseudomonas wisconsinensis; lipase modulator; expression system;
KW   enzyme production; detergent.
OS   Pseudomonas wisconsinensis.
PN   WO9713847-A1.
PD   17-APR-1997.
PF   14-OCT-1996; BE0109.
PR   12-OCT-1995; BE-000851.
PA   (GEMV ) GENENCOR INT INC.
PI   Andre C, Charmoille L, Cornelis P, Dhaese P, Hazbon MH;
DR   WPI; 97-235883/21.
DR   N-PSDB; T73345.
PT   Expression system containing regulatory sequences from P.
PT   wisconsinensis - and optionally modulator and GPW protein sequences,
PT   especially for lipase production for use e.g. in detergents
PS   Claim 23; Pages 36-38; 74pp; French.
CC   This sequence, isolated from Pseudomonas wisconsinensis, is a lipase
CC   modulator. A claimed expression system for production of an enzyme,
CC   especially a Pseudomonas lipase, comprises at least sequences for a
CC   promoter, signal sequence, mature enzyme and a terminator. The
CC   transformed cells are used to produce enzymes, especially lipase for use
CC   in detergents. Other enzymes are useful in food, pharmaceutical and
CC   chemical industries. The system provides effective secretion of large
CC   quantities of enzyme into the culture medium. GPW seems to assist
CC   expression of enzyme and to protect it (and cells) against oxygen
CC   radicals, also to repair cell membranes damaged by oxygen radicals.
SQ   Sequence    352 AA;

Query Match           3.2%;  Score 7;  DB 1;  Length 352;
Best Local Similarity  100.0%;  Pred. No. 30;
Matches    7;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    204 YTLDSLA 210
          |||||||
Db    203 YTLDSLA 209

RESULT 6
W55695
ID   W55695 standard; Protein; 398 AA.
AC   W55695;
```

DT 07-JUL-1998 (first entry)
DE H. pylori ORF 13ae10610_859692_c2_32 cytoplasmic protein.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI: 97-503122/46.
DR N-PSDB: V25104.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 14,96; Pages 957-958; 1145pp; English.
CC This sequence is a H. pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds. The
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence   398 AA;

Query Match   3.2%;   Score 7; DB 1;   Length 398;
Best Local Similarity   100.0%;   Pred. No. 34;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 128 RNGVTLN 134
        |||||||
Db 210 RNGVTLN 216

RESULT   7
R44434
ID R44434 standard; Protein; 416 AA.
AC R44434;
DT 06-JUN-1994 (first entry)
DE Subtilisin-like serine protease PEPD from Aspergillus niger.
KW PEPD; serine protease; subtilisin; filamentous fungus;
KW protease-deficient yeast strain.
OS Aspergillus niger (strain N400).
PN AU9336959-A.
PD 21-OCT-1993.
PF 15-APR-1993; 036959.
PR 15-APR-1992; EP-810281.
PR 12-MAR-1993; GB-005097.
PA (CIBA) CIBA GEIGY AG.
PI Buxton F;
DR WPI: 93-386961/49.
DR N-PSDB: Q52536.
PT New mutant subtilisin-type serine protease(s) derived from

PT Aspergillus niger - used to transform hosts to confer reduced
PT protease activity, for improved yields in protein expression
PS Claim 17; Page 70-72; 85pp; English.
CC The pepD gene was isolated from A.niger N400 and sequenced. The gene
CC codes for a subtilisin-type protease PEPD. Engineered A.niger
CC mutants which are unable to produce functional PEPD can be used for
CC increased recombinant production of heterologous proteins.
SQ Sequence   416 AA;

Query Match   3.2%;   Score 7; DB 1;   Length 416;
Best Local Similarity   100.0%;   Pred. No. 35;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 139 LPAGIER 145
        |||||||
Db   75 LPAGIER 81

RESULT   8
R84345
ID R84345 standard; Protein; 764 AA.
AC R84345;
DT 25-MAR-1996 (first entry)
DE Melon aconitase encoded by cDNA clone 16.
KW Arabidopsis thaliana; aconitase; exon; intron; probe; melon; zea mais;
KW Cucumis melo; maize; plant metabolism; Krebs cycle; glyoxylate cycle;
KW citrate; acetyl CoA; catabolism; polysaccharide; lipase; chimeric;
KW resistance marker; hormone; enzyme; ss.
OS Cucumis melo.
PN WO9520046-A1.
PD 27-JUL-1995.
PF 25-JAN-1995; E00263.
PR 25-JAN-1994; FR-000787.
PA (BIOC-) BIOCEM SA.
PI Alric M, Perez P, Peyret P;
DR WPI: 95-269459/35.
DR N-PSDB: T02367.
PT New plant aconitase, its fragments and related nucleic acid - also
PT chimeric genes, transgenic plants, antibodies etc., used to modify
PT plant metabolism by regulating carboxylic acid prodn.
PS Example 4; Fig 4; 122pp; French.
CC The amino acid sequence of the melon aconitase encoded by the melon
CC aconitase gene clone 16. The gene was isolated by screening a melon cDNA
CC library with antibodies raised against the purified aconitase from melon
CC seeds. The melon seed aconitase was compared to the potato (S.tuberosum)
CC mitochondrial aconitase and good homology in the N-terminal sequence was
CC observed. A 2.2 kb EcoRI fragment of the melon aconitase cDNA was used as
CC a probe to isolate the sequence of the Arabidopsis thaliana aconitase
CC cDNA sequence (T02366). The A.thaliana gene sequence was then used to
CC obtain a partial gene sequence of the maize aconitase gene (T02365). The
CC aconidase genes can be used to modify plant metabolism by overexpression
CC of aconitase. This leads to overproduction of acids in the Krebs and
CC glyoxylate cycles, esp. citrate. Fragments of the genes can be used to
CC inhibit the expression of aconitase, resulting in overproduction of
CC acetyl CoA, and alterations of metabolism/catabolism of polysaccharides,
CC lipases, and N cpds. Chimeric genes contg. the aconitase gene or its
CC promoter, are used to provide controlled expression e.g. during
CC development, or to express heterologous enzymes, desired traits,
CC resistance markers, hormones, etc., in plants.
SQ Sequence   764 AA;

Query Match   3.2%;   Score 7; DB 1;   Length 764;
Best Local Similarity   100.0%;   Pred. No. 62;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 126 KLRNGVT 132
        |||||||
Db 120 KLRNGVT 126

RESULT   9

OM of: US-09-252-691-7056 to: GenEmbl:*   out_format : pfs
Date: Jun 10, 2000 12:46 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US0925269l/runat_05062000_101650_595/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=0.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09252691
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: GenEmbl:*
Database sequences: 882769
Database length:-48639572 9
Search time (sec): 563.060000

```
gb_bal:BATRYOPEA    -  175.50  237.34  3.6e-05   8392  ! Z19055 B.aphidicola tryptop
gb_bal:MHX99740     -  166.00  221.38  0.0003   10414  ! X99740 Mycoplasma hominis
gb_bal:RP282292     -  163.00  246.39  1.1e-05    333  ! 282292 R.prowazekii genomic
```

seq_name: gb_ba2:AE000213

seq_documentation_block:
```
LOCUS       AE000213   10959 bp      DNA            BCT    12-NOV-1998
DEFINITION  Escherichia coli K-12 MG1655 section 103 of 400 of the complete
            genome.
ACCESSION   AE000213 U00096
VERSION     AE000213.1  GI:1787371
KEYWORDS    .
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 10959)
AUTHORS     Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
TITLE       The complete genome sequence of Escherichia coli K-12
JOURNAL     Science 277 (5331), 1453-1474 (1997)
MEDLINE     97426617
REFERENCE   2  (bases 1 to 10959)
AUTHORS     Blattner,F.R.
TITLE       Direct Submission
JOURNAL     Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10959)
AUTHORS     Blattner,F.R.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10959)
AUTHORS     Plunkett,G. III.
TITLE       Direct Submission
JOURNAL     Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K-12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K-12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES            Location/Qualifiers
     source         1..10959
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
gb_ba2:AE000213	-	983.00	1434.19	7.9e-72	10959	! AE000213 Escherichia coli K-1
gb_bal:D90748	-	983.00	1431.50	1.1e-71	15007	! D90748 Escherichia coli genom
gb_bal:D90749	-	983.00	1428.92	1.5e-71	20284	! D90749 Escherichia coli genom
gb_htg7:AC022161	+	850.00	1212.68	1.7e-59	180993	! AC022161 Homo sapiens chromo
gb_htg7:AC009127	+	824.00	1173.81	2.5e-57	186591	! AC009127 Homo sapiens chromo
gb_htg6:AC008911	+	696.50	982.80	1.1e-46	226805	! AC008911 Homo sapiens chromo
gb_ba2:U32752	+	598.00	861.81	6.0e-40	11833	! U32752 Haemophilus influenzae
gb_ba2:SLE250721	+	475.50	686.49	3.5e-30	5476	! AJ250721 Synechococcus leopoli
gb_bal:D90899	+	474.50	657.65	1.4e-28	133859	! D90899 Synechocystis sp. PCC
gb_bal:SYCSLRA	+	298.00	396.13	5.2e-14	125469	! D63999 Synechocystis sp. PCC
gb_ba2:AE001942	+	272.00	379.14	4.6e-13	9999	! AE001942 Deinococcus radiodura
gb_ba2:AE001708	-	264.50	362.27	4.6e-12	19531	! AE001708 Thermotoga maritima
gb_ba2:AE000741	-	259.00	356.32	8.6e-12	15077	! AE000741 Aquifex aeolicus sec
gb_bal:BACDIA	+	253.00	342.05	5.4e-11	28206	! L09228 Bacillus subtilis spoV
gb_bal:BSUB0013	+	253.00	324.52	2.1e-10	218470	! 299116 Bacillus subtilis con
gb_ba2:ECU18111	+	245.00	348.24	2.4e-11	3415	! U18111 Escherichia coli region
gb_ba2:AE000225	+	245.00	337.59	9.5e-11	11850	! AE000225 Escherichia coli K-1
gb_bal:D90764	+	245.00	334.72	1.4e-10	16573	! D90764 E.coli genomic DNA, Kc
gb_bal:D90765	+	245.00	333.57	1.6e-10	18962	! D90765 E.coli genomic DNA, Kc
gb_ba2:U32799	+	244.50	337.41	9.7e-11	11093	! U32799 Haemophilus influenzae
gb_htg7:AC022162	-	236.50	302.10	3.0e-09	171111	! AC022162 Homo sapiens chromo
gb_htg7:MCI125	-	233.50	310.67	3.0e-09	37432	! 298268 Mycobacterium tubercul
gb_htg11:AC021060	+	224.00	280.11	1.5e-07	255800	! AL109848 Streptomyces coelic
gb_bal:SCI51	-	218.50	287.67	6.7e-08	40745	! 295117 Mycobacterium leprae c
gb_bal:MCB1351	-	217.50	286.57	6.6e-08	38936	! U00021 Mycobacterium leprae c
gb_bal:U00021	+	214.00	281.32	1.3e-07	39193	! U00021 Mycobacterium leprae c
gb_bal:RPXX03	+	209.50	257.83	2.6e-06	279110	! AJ235272 Rickettsia prowazek
gb_pat:AR007490	+	206.00	291.26	3.6e-08	3061	! AR007490 Sequence 1 from paten
gb_ba2:AE000308	-	206.00	279.59	1.6e-07	11961	! AE000308 Escherichia coli K-1
gb_bal:ECOHU49	-	206.00	269.45	5.9e-07	39149	! U00008 centisome 49 region of
gb_bal:BSUB0016	-	205.50	254.10	4.2e-06	215640	! 299119 Bacillus subtilis com
gb_ba2:AF008220	+	205.50	253.92	4.3e-06	220060	! AF008220 Bacillus subtilis r
gb_ba2:AE000475	+	201.50	273.86	3.4e-07	10703	! AE000475 Escherichia coli K-1
gb_ba2:ECOUW89	+	201.50	249.89	7.3e-06	176195	! U00006 E. coli chromosomal r
gb_htg6:AC010542	+	201.50	247.51	9.9e-06	232632	! AC010542 Homo sapiens chromo
gb_htg10:AC016580	+	201.50	245.58	1.3e-05	291282	! AC016580 Homo sapiens chromo
gb_ba2:AE000695	+	200.50	269.16	6.1e-07	15585	! AE000695 Aquifex aeolicus sec
gb_ba2:AF052507	+	193.00	271.96	4.3e-07	3061	! AF052507 Pseudomonas putida ac
gb_bal:ECBICRES	-	187.00	266.60	8.5e-07	2020	! X63703 E.coli DNA involved in
gb_ba2:U32804	-	186.00	249.53	7.6e-06	12481	! U32804 Haemophilus influenzae
gb_ba2:AE001667	+	178.50	234.19	5.4e-05	20386	! AE001667 Chlamydia pneumoniae
gb_htg5:AC018634	-	176.50	207.78	0.0016	315277	! AC018634 Homo sapiens chromo

```
                    /organism="Escherichia coli"
                    /strain="K-12"
                    /sub_strain="MG1655"
                    /db_xref="taxon:562"
promoter        13..40
                    /note="factor Sigma70; predicted +1 start at 1184959"
protein_bind    119..136
                    /note="central position to predicted promoter:82"
                    /bound_moiety="PhoB predicted site"
gene            155..1381
                    /gene="pepT"
                    /note="b1127"
CDS             155..1381
                    /gene="pepT"
                    /EC_number="3.4.11.-"
                    /function="putative enzyme; Not classified"
                    /note="o408; 100 pct identical to 42 aa fragment
                    PEPT_ECOLISW: P29745; 93 pct identical to 407 residues of
                    409 aa PEPT_SALTY SW: P26311"
                    /codon_start=1
                    /transl_table=11
                    /product="putative peptidase T"
                    /protein_id="AAC74211.1"
                    /db_xref="GI:1787372"
                    /db_xref="PID:g1787372"
                    /translation="MDKLLERFLNTVVSLDDTQSKAGVRQVPSTEGQWKLLHLLKEQLEE
                    MGLINVTLSEKGTLMATLPANVPGDIPAIGFISHVDTSPDCSGKNVNPQIVENYRGGD
                    IALGIGDEVLSPVMFPVLHQLLGQTLITTDGKTLLGADDKAGIAEIMTALAVLQQKKI
                    PHGDIRVAFTPDEEVKGAKHFDVDAFDARWAYTVDGGVGELEFENFNAASVNIKIV
                    GNNVHPGTAKGVMVNALSLAARIHAEVPADESPEMTEGYEGFYHLASMKGTVERADMH
                    YIIRDFDRKQFEARKRRMMEIAKKVKGLHPDCYIELVIEDSYYNMREKVVEHPHILD
                    IAQQAMRDCDIEPELKPIRGGTDGAQLSFMGLPCPNLFTGGYNYHGKHEFVTLEGMEK
                    AVQVIVRIAELTAQRK"
gene            complement(1430..2560)
                    /gene="ycfD"
                    /note="b1128"
CDS             complement(1430..2560)
                    /gene="ycfD"
                    /function="orf; Unknown"
                    /note="f376; 100 pct identical to fragment YCFD_ECOLI
                    SW:P27431 but has 3 additional N-terminal aa and 93
                    additional C-terminal aa"
                    /codon_start=1
                    /transl_table=11
                    /product="orf, hypothetical protein"
                    /protein_id="AAC74212.1"
                    /db_xref="GI:1787373"
                    /db_xref="PID:g1787373"
                    /translation="MLNMEYQLTLNWPDFLERHWQKRPVVLKRGFNNFIDPISPDELA
                    GLAMESEVDSRLVSHQDGWKQVSHGPFESYDHLGETNWSLLVQAVNHWHEPTAALMRP
                    FRELPDWRIDDLMISFSVPGGVGPHLDQVFIIQGTGRRWRVGEKLQMKQHCPHP
                    DLLQVDPFEAIIDEELEPGDILYIPPGFPHEGYALENAMNYSVGFRAPNTRELISGFA
                    DYVLQRELGGNYYSDPDVPPRAHPADVLFQEMDKLREMMLELINQPEHFKQWFGEFIS
                    ALDALASNIALTAENFGDALEDPSFLAMLAALVNSGYWFFEG"
promoter        complement(2583..2613)
                    /note="factor Sigma70; predicted +1 start at 1187488"
gene            complement(2627..4087)
                    /gene="phoQ"
                    /note="b1129"
CDS             complement(2627..4087)
                    /gene="phoQ"
                    /EC_number="2.7.3.-"
                    /function="enzyme; Global regulatory functions"
                    /note="f486; 99 pct identical to PHOQ_ECOLI SW: P23837"
                    /codon_start=1
                    /transl_table=11
                    /product="sensor protein PhoQ"
                    /protein_id="AAC74213.1"
                    /db_xref="GI:1787374"
                    /db_xref="PID:g1787374"
                    /translation="MKKLLRLFFPLSLRVRFLLATAAVVLVLSLAYGMVALIGYSVSF
                    DKTTFRLLRGESNLFYTLAKWENNKLHVELPENIDKQSPTMTLIYDENGQLLWAQRDV

                    PWLMKMIQPDWLKSNGFHEIEADVNDTSLLSGDHSIQQOLQEVREDDDAEMTHSVA
                    VNVYPATSRMPLTIVVVDTIPVELKSSYMWSWFIYVLSANLLLVIPLLWVAAWSL
                    RPIEALAKEVRELEEHNRELLNPATTRELTSLVRNLNRLLKSERERYDKYRTTLTDLT
                    HSLKTPLAVLQSTLRSLRSERKMSVSDAEPVMLEQISRISQOIGYYLHRASMRGGTLLS
                    RELHPVAPLLDNLTSALNKVYQRKGVNISLDISPEISFVGEQNDFVEVMGNVLDNACK
                    YCLEFVEISARQTDEHLYIVVEDDGPGIPLSKREVIFDRGQRVDTLRPGQGVGLAVAR
                    EITEQYEGKIVAGESMLGGARMEVIFGRQHSAPKDE"
gene            complement(4087..4758)
                    /gene="phoP"
                    /note="b1130"
CDS             complement(4087..4758)
                    /gene="phoP"
                    /function="regulator; Global regulatory functions"
                    /note="f223; 100 pct identical to PHOP_ECOLI SW: P23836"
                    /codon_start=1
                    /transl_table=11
                    /product="transcriptional regulatory protein"
                    /protein_id="AAC74214.1"
                    /db_xref="GI:1787375"
                    /db_xref="PID:g1787375"
                    /translation="MRVLVVEDNALLRHHLKVQIQDAGHQVDDAEDAKEADYLNEHI
                    PDIAIVDLGLPDEDGLSLIRRWRSNDVSLPFILVLTARESWQDKVEVLSAGADDYTKP
                    FHIEEVMARMQALMRRNSGLASQVISLPPFQVDLSRRELSINDEVIKLTAFEYTIMET
                    LIRNNGKVVSKDSLMLQLYPDAELRESHTIDVLMGRLRKKIQAQYPQEVITTVRGQGY
                    LFELR"
promoter        complement(4257..4287)
                    /gene="phoP"
                    /note="factor Sigma70; predicted +1 start at 1189162"
gene            complement(4927..6297)
                    /gene="purB"
                    /note="b1131"
CDS             complement(4927..6297)
                    /gene="purB"
                    /EC_number="4.3.2.2"
                    /function="enzyme; Purine ribonucleotide biosynthesis"
                    /note="f456; 99 pct identical to PUR8_ECOLI SW: P25739"
                    /codon_start=1
                    /transl_table=11
                    /product="adenylosuccinate lyase"
                    /protein_id="AAC74215.1"
                    /db_xref="GI:1787376"
                    /db_xref="PID:g1787376"
                    /translation="MELSSLTAVSPVDGRYGDKVSALRGIFSEYGLLKFRVQVEVRWL
                    QKLAHAAIKEVPAFAADAIGYLDAIVASFSEEDAARIKTIERTNHDVKAVEYFLKE
                    KVAEIPELHAVSEFIHFACTSEDINNLSHALMLKTARDEVILPYWRQLIDGIKDLAVQ
                    YRDIPLLSRTHGQPATPSTIGKEMANVAYRMEROYRQLNQVEILGKINGAVGNYNAHI
                    AAYPEVDWHQFSEEFVTSLGIQWNPYTTQEPHDYIAELFDCVARFNTILIDFDRDVW
                    GYIALNHFKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNAVLQHLASKLPVSRWQ
                    RDLIDSTVLRNLGVGIAYALIAYQSTLKGVSKLEVNRDHLLDELDHNWEVLAEPIQTV
                    MRRYGIEKPYEKLKELTRGKRVDAEGMKQFIDGLALPEEKARLKAMTPANYIGRAIT
                    MVDELK"
gene            complement(6301..6942)
                    /gene="ycfC"
                    /note="b1132"
CDS             complement(6301..6942)
                    /gene="ycfC"
                    /function="orf; Unknown"
                    /note="f213; 99 pct identical to YCFC_ECOLI SW: P25746"
                    /codon_start=1
                    /transl_table=11
                    /product="orf, hypothetical protein"
```

alignment_scores:
   Quality: 983.00   Length: 218
   Ratio: 4.795   Gaps: 0
Percent Similarity: 94.037 Percent Identity: 85.780

alignment_block:
US-09-252-691-7056 x AE000213/rev
Align seg 1/1 to reverse of: AE000213 from: 1 to: 10959

 1 AlaIleMetArgGlnLeuIleThrProGluAsnThrMetThrLysThrSe 17

```
              ||||||||||||||||::|||        |||||||||||   |||||||
9268  GCTATAATGCGGCAATTCATAATCTCTGAAAATACCATGCAAAAACTTC  9219

 17   rPheArgGlysHisArgValGluArgPheSerSerArgGlnAlaThrArgA  34
      ||||::|||::||::|||||||||:::||||||::::|||:||||||||||
9218  TTTTAGAAATCACCAGGTAAGCGATTCAGCTCGCAACGTTCTACCAGGC  9169

 34   rgThrProGluProGlnProThrArgValIleLeuPheAsnLysProTyr  50
      ||  |||||  ||||||||||||||||::||||||||::||||||||||||
9168  GTAAACCTGAAAACCAGCCCACCGGTGTGATCCTGTTCAATAAACCCTAC  9119

 51   AspValLeuProGlnPheThrAspGluAlaGlyAlaGlyArgSerThrLeuLysAs  67
      ||||||||||||||||::||||||||||::|||||:::||||||::|||||
9118  GATGTTCTTCCGCAGTTCACCGATGAAGCCGGACGCAAAACATTAAAAGA  9069

 67   pPheIleProValGlnGlyValTyrAlaAlaGlyValArgLeuAspArgAspS  84
      :||||||||||||||||||||||||||||||||||:::||||||||||||
9068  ATTCATCCCGGTTCAGGGTGTTTATGCAGCAGGTCGCCTTGACCGCGATA  9019

 84   erGluGlyLeuLeuValLeuThrAsnAspPheGlyValLeuGlnAlaAlaArgLeu  100
      ||||||||||||||||||:::|||::||||||||||||||||||||||||
9018  GCGAAGGGTTGCTGGTGCTGACCAATAACGGCCGTTGCCAGGCGCGTTTA  8969

101   ThrGlnProGlyLysArgThrrGlyTyrLysIleTyrTyrValGlnValGluGl  117
      |||||||||||||||::|||||||||||||||||||||||||||||||||
8968  ACCCAGCCCGGTAAACGCACCGGAAAAATCTATTATGTGCAGGTGGAAGG  8919

117   yGluProAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuA  134
      ||  ::    |||||||    ||||||||||||||||||||||||||||||
8918  TATTCCCACACAAGACGCACTTGAAGCCTTGCCCAATGGCGTAACCTTAA  8869

134   snAspGlyProThrLeuProAlaGlyIleLeuGluArgValAsnGluProGlu  150
      ||||||||||||||||||||  |||  ::|||||||||||||||||||||
8868  ATGATGGCCCTACCCTGCCCGCCGGCGGAGCTGGTTGACGAACCCGCG  8819

151   TrpLeuTrpProArgAsnProProIleArgGluArgGlyLysSerIleProThr  167
      ||||||||||||||||||||||||||||||||::|||||||||||||:||
8818  TGGTTATGGCCGCGGAATCCACCAATCCGTGAACGCAAAAGTATTCCCAC  8769

167   rSerTrpLeuLysIleThrLeuTrpTyrGluGlyArgAsnArgGlnValArgA  184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
8768  CAGCTGGCTGGAGATCACCTTATGGAAGGACGTAATCGCCAGGTGCCGCC  8719

184   rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla  200
      |||||||||||||||||||||||||||||::||||||||||||||||||
8718  GCATGACCGCCCATGTTGGCTTCCCCACGCTGCGACTGATTCGCTATGCG  8669

201   MetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgArgAspVa  217
      ||||||||||||||:::||||||||||||||||:::||||||||||::||
8668  ATGGGTAGCTATACTCTTTGGATAATCTTGCCAATGGTGAATGGCGAGAAGT  8619

217   lThr  218
      ||||
8618  GACA  8615
```

q_name: gb_bal:D90748

```

```
RLLDPLYFEVLLHSLNMALIATACLVLGYPFAWFLAKLPHKVRPLLLFLLIVPFWTN
SLIRIYGLKIFLSTKGYLNEFLLWLGVIDTPIRIMFTPSAVIIGLVYILLPFMVMPLY
SSIERLDKPLEAARDLGASKLQTFIRIIIPLIMPGIIAGCLLVMLPAMGLFYVSDLM
GGAKNLLI"
```

gene    complement(692..1828)
        /gene="potA"
CDS     complement(692..1828)
        /gene="potA"
        /note="ORF_ID:o2391#1; similar to PIR Accession Number
        A40840"
        /codon_start=1
        /transl_table=11
        /product="spermidine/putrescine transport protein A"
        /protein_id="BAA35948.1"
        /db_xref="GI:1651555"
        /translation="MGQSKKLNKQPSSLSPLVQLAGIRKCFDGKEVIPQLDLTINNGE
FLILLGPSGCGKTTVLRLIAGLETVDSGRIMLDNEDITHVPAENRYVNTVFQSYALFP
HMTVFENVAFGLRMQKTPAAEITPRVMEALRMVQLETFAQRKPHQLSGGQQQRVAIAR
AVVNKPRLLLLDESLSALDYKLRQMONELKALQRKLGITFVFVTHDQEEALMSDRI
VVMRDGRIEQDGTPREIYEPKNLFVAGFIGEINMFNATVIERLDEQRVRANVEGREC
NIYVNFAVEPGQKLHVLLRPEDLRVEEINDDNHAEGLIGYVRERNYKGMTLESVVELE
NGKMVMVSEFFNEDDPDFDHSLDQKMAINWVESWEVVLADEEHH"

gene    2078..3304
        /gene="pepT"
CDS     2078..3304
        /gene="pepT"
        /note="ORF_ID:o2392#1; similar to SwissProt Accession
        Number P29745"
        /codon_start=1
        /transl_table=11
        /product="Peptidase T (EC 3.4.11.-) (aminotripeptidase)
        (tripeptidase)"
        /protein_id="BAA35949.1"
        /db_xref="GI:1651556"
        /translation="MDKLLERFLNYVSLDTQSKAGVRQVPSTEGQWKLLHLLKEQLEE
MGLINVTLSEKGTLMATLPANVPGDIPAIGFISHVDTSPDCSGKNVNPQIVENYRGGD
IALGIGDEVLSPVMFPVLHQLLGQTLITTDGKTLLGADDKAGIAEIMTALAVLQQKKI
PHGDIRVAFTPDEEVGKGAKHFDVDAFDARWAYTVDGGGVGELEFENFNAASVNIKIV
GNNVHPGTAKGVMVNALSLAARIHAEVPADESPEMTEGYEGFYLASMGTVERADMH
YIIRDFDRKQFEARRKRKMEIAKKVGKGLHPDCYIELVIEDSYINMREKVVEHPHILD
IAQQAMRDCDIEPELKPIRGGTDGAQLSFMGLPCPNLFTGGYNYHGKHEFVTLEGMEK
AVQVIVRIAELTAQRK"

gene    complement(3353..6010)
        /gene="phoQ"
CDS     complement(3353..4474)
        /gene="phoQ"
        /note="ORF_ID:o2393#3; similar to PIR Accession Number
        D41966"
        /codon_start=1
        /transl_table=11
        /product="Orf 2 downstream of phoQ."
        /protein_id="BAA35950.1"
        /db_xref="GI:4062695"
        /translation="MEYQLTLNWPDFLERHWQKRPVVLKRGFNNFIDPISPDELAGLA
MESEVDSRLJVSHQDGKWQVSHGPPESYDHLGETNWSLLVQAVNWHEPTAALMRPFRE
LPDWRIDDLMISFSVPGGGVGPHLDQYDVFIIQGTGRRRWRVGEKLQMKQHCPHPDLL
QVDPFEAIIDEELEPGDILYIPPGFPHEGYALENAMNYSVGFRAPNTRELISGFADYV
LQRELGGNYYSDPDVPPRAHPADVLPQEMDKLREMMLELINQPEHFKQWGEFISQSR
HELDIAPPEPPYQPDEIYDALKQGEVLVRLGGLRVLRIGDDVYANGEKIDSPHRPALD
ALASNIALTAENFGDALEDPSFLAMLAALVNSGYWFFEG"

gene    complement(4550..6010)
        /gene="phoQ"
CDS     complement(4550..6010)
        /gene="phoQ"
        /note="ORF_ID:o2394#; similar to PIR Accession Number
        B41966"
        /codon_start=1
        /transl_table=11
        /product="Virulence membrane protein phoQ."
        /protein_id="BAA35951.1"
        /db_xref="GI:1651557"
        /translation="MKKLLRLFFPLSLRVRFLLATAAVVLVLSLAYGMVALIGYSVSF
DKTFFRLLRGESNLFYTLAKWENNKLHVELPENIDKQSPTMLIYDENGQLLWAGRDV
PWLMKMIQPDWLKSNGFHEIEADVNDTSLLLSGDHSIQQQLQEVREDDDAEMTHSVA
VNYPATSRMPKLTIVVVDTIPVELKSSTMVWSWFIYVLSANLLLVIPLLWVAAWSL"

```
RPIEALAKEVRELEHNRELLNPATTRELTSLVRNLRLLKSERERYDKYRTTLIDLT
HSLKTPLAVLQSTLRSLRSEKMSVSDAEPVMLEQISRISQOIGYIHRASMRGGTLLS
RELHPVAPLLDNLTSALNKVYQRKGVNISLDISPEISFVGEQNDFVEVMGNVLDNACK
YCLEFVEISARQTDEHLYIVVEDDGPGIPLSKREVIFDRGQRVDTLRPGGVGLAVAR
EITEQYEGKIVAGESMLGGARMEVIFGRQHSAPKDE"
```

gene   complement(6010..6681)
        /gene="phoP"
CDS   complement(6010..6681)
        /gene="phoP"
        /note="ORF_ID:o2401#; similar to SwissProt Accession
        Number P23836"
        /codon_start=1
        /transl_table=11
        /product="Transcriptional regulatory protein PhoP."
        /protein_id="BAA35952.1"
        /db_xref="GI:1651558"
        /translation="MRVLVVEDNALLRHHLKVQIODAGHQVDDAEDAKEADYLNEHI
PDIAVDLGLPDEDGLSLIRRWRSNDVSLPILVLTARESWQDKVEVLSAGADDYTKP
FHIEVMARMQALMRRNSGLASQVISLPPFQVDLSRRELSINDEVIKLTAFEYIMET
LIRNNGKVVSKDSLMLQLYPDAELRESHTIDVLMGRLRKKIQAQYPQEVITTVRGQGY
LFELR"

gene   complement(6850..8220)
        /gene="purB"
CDS   complement(6850..8220)
        /gene="purB"
        /note="ORF_ID:o2402#2; similar to PIR Accession Number
        S19212"
        /codon_start=1
        /transl_table=11
        /product="Adenylosuccinate lyase (EC 4.3.2.2)"
        /protein_id="BAA35953.1"
        /db_xref="GI:1651559"
        /translation="MELSSLTAVSPVDGRYGDKVSALRGIFSEYGLLKFRVQVEVRWL
QKLAAHAAIKEVPAFAADAIGYLDAIVASFSEEDAARIKIIERTNHDVKAVEYFLKE
KVAEIPELHAVSEFIHFACTSEDINNLSHALMLKTARDEVILPYWRQLIDGIKDLAVQ
YRDIPLLSRTHGQPATPSTIGKEMANVAYRMEROYRQLNOVEILGKINGAVGNYNAHI
AAYPWDHQFSEEVTSLGIQWNPYTTQIEPHDYIAELFDCVARFNTILIDFDRDVW
GYIALNFHKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNAVLQHLASKLPVSRWQ
RDLTDSTVLRNLGVIGYALIAYQSTLKGVSKLEVNRDHLLDELDHNWEVLAEPIQTV"

```
alignment_scores:
        Quality:       983.00        Length:      218
        Ratio:           4.795        Gaps:          0
Percent Similarity:    94.037     Percent Identity:   85.780

alignment_block:
US-09-252-691-7056 x D90748/rev

Align seg 1/1 to reverse of: D90748    from: 1    to: 15007

    1 AlaIleMetArgGlnLeuIleThrProGlnAsnThrMetThrLysThrSe   17
      ||||||||||||||||||||||||||||||||||| |||| |||||||||
11191 GCTATAATGCGGCAATTCATATCTCTGAAATACCATGCAAAAAACTTC     11142

   17 rPheArgLysHisArgGlnArgPheSerArgGlnAlaThrArgA        34
      |||||||   || || |||||||   ||||||||| ::|||||||
11141 TTTTAGAAATACCAGGTTAAGCGCAATTCAGCTCGCAACGTCTCCACCAGGC  11092

   34 rgThrProGluProGlnProThrArgValIleLeuPheAsnLysProTyr   50
      ||  |||||  |||||||||||||||||||||||||||||||||||||||||
11091 GTAAACCTGAAAACCAGCCCACCGGGTCGATCCTGTTCAATAAACCCTAC    11042

   51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs   67
      ||||||||||||||||||||||||| || |||||||| | ::||||||||
11041 GATGTGTCTTCCCCAGTTCACCGATGAGGCCGGACGCAAACATTAAAAGA    10992

   67 pPheIleProValGlnGlyValTyrAlaAlaAlaGlyArgLeuAspArgAspS  84
      :||||||||:||::||||||||||:||||||||||:||:::||||||||:
10991 ATTCATCCCGGTTCAGGGTGTTTATGCAGCAGGTCGCCTTGACCGGCGCATA  10942

   84 erGluGlyLeuLeuValLeuThrAsnAspValValLeuGlnAlaArgLeu    100
      |||||||||||||||||| |||| :::|| ||||||||||||||||||||
10941 GCGAAGGGTTGCTGGTGCTGACCAATAACGGCGCCGGTTGCAGGCGCGTTTA  10892
```

```
  101 ThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGl 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
10891 ACCCAGCCGGGTAAACGCACCGGAAAAATCTATTATGTGCAGGTGGAAGG 10842

  117 yGluProAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuA 134
      ||||||  :::  ||  : ::::||  ||||||||||||||||||||
10841 TATTCCCACACAAGACGCACTTGAAGCCTTGCGCAATGGCGTAACCTTAA 10792

  134 snAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProGlu 150
      |||||||||||||||||||||||| |||| :::||||||||||||||||
10791 ATGATGGCCCTACCCTGCCCGCCGGCGGAGCTGGTTGACGAACCCGCG 10742

  151 TrpLeuTrpProArgAsnProProIleArgGluArgGlyLysSerIleProThr 167
      |||||||||||||||||||||||||||||||||||||||||||||||||
10741 TGGTTATGGCCGCGGAACCCACCAATTCGTGAACGCAAAAGTATTCCCAC 10692

  167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
      ||||||||||||||||||||||||||||||||||||||||||||||||
10691 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGCC 10642

  184 rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
10641 GCATGACCGCCCATGTTGGCTTCCCCACCGCTGCGACTGATTCGCTATGCG 10592

  201 MetGlySerTyrThrLeuAspAspSerLeuAlaAsnGlyGluTrpArgAspVa 217
      ||||||:|||||::|||||||||||||||:|||||||||||||:::|||
10591 ATGGGTAGTTACTCTTTGGATAATCTTGCCAATGTGAATGGGGAGAAGT 10542

  217 lThr 218
      |||
10541 GACA 10538
```

COMMENT   Collaboration Information:
          Project:
          The Japan E.coli genome DNA sequencing project
          Group:
          The Japan E.coli genome DNA sequencing group
          Members: (1995.4 - 1996.3)
          Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
          Horiuchi,T., Ikemoto,K. Inada,T., Isono,K., Isono,S.,
          Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
          Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
          Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
          Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
          Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
          Yamamoto,Y. and Yano,M.
          Headed by:
          Name: Takashi Horiuchi
          Address: National Institute of Basic Biology, Okazaki, 444, Japan
          E-mail: kishori@nibb.ac.jp
          Information operator:
          Name: Hirotada Mori
          Address: NARA Institute of Science and Technology,
                   Ikoma, 630-01, Japan
          E-mail: hmori@gtc.aist-nara.ac.jp
          URL:
          The Japan E. coli genome database
          http:bsw3.aist-nara.ac.jp.
FEATURES   Location/Qualifiers
     source   1..20284
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /clone="Kohara clone #240"
                     /map="25.7-26.1 min"
                     /note="Nucleotide position 1191151-1211434 from the
                     initiation site of ThrA (0 min.).; This clone is from
                     Kohara lambda miniset library."
     gene   complement(1..327)
                     /gene="phoQ"
     CDS   complement(<1..327)
                     /gene="phoQ"
                     /note="ORF_ID:o23944; similar to PIR Accession Number
                     B41966"
                     /codon_start=1
                     /transl_table=11
                     /product="virulence membrane protein phoQ."
                     /protein_id="BAA35960.1"
                     /db_xref="GI:1651563"
                     /translation="MKKLLRLFFPLSLRVRFLLATAAVVLVLSLAYGMVALIGYSVSF
                     DKTFRLLRGESNLFYTLAKWENNKLHVELPENIDKQSPTMTLIYDENGQLLWAQRDV
                     PWLMKMI"
     gene   complement(327..998)
                     /gene="phoP"
     CDS   complement(327..998)
                     /gene="phoP"
                     /note="ORF_ID:o240#1; similar to SwissProt Accession
                     Number P23836"
                     /codon_start=1
                     /transl_table=11
                     /product="Transcriptional regulatory protein PhoP."
                     /protein_id="BAA35961.1"
                     /db_xref="GI:1651564"
                     /translation="MRVLVVEDNALLRHHLKVQIQDAGHQVDDAEPEADYLNEHI
                     PDIAIVDLGLPDEDGLSLIRRWRSNDVSLPILVLTARESWQDKVEVLSAGADDYVTKP
                     FHIEEVMARMQALMRRNSGLASQVISLPPFQVDLSRRELINDEVIKLTAFEYIMET
                     LIRNNGKVVSKDSLMLQLYPDAELRESHTIDVLMGRLRRKIQAQYPQEVITTVRGQGY
                     LFELR"
     gene   complement(1167..2537)
                     /gene="purB"
     CDS   complement(1167..2537)
                     /gene="purB"
                     /note="ORF_ID:o240#2; similar to PIR Accession Number

seq_name: gb_bal:D90749

seq_documentation_block:
LOCUS   D90749   20284 bp   DNA   BCT   07-FEB-1999
DEFINITION Escherichia coli genomic DNA. (25.7 - 26.1 min).
ACCESSION   D90749 AB001340
VERSION   D90749.1 GI:1651562
KEYWORDS   Complete and shotgun sequencing; phoQ; phoP; purB; ycfC; ycfB;
            lcdA; lcd; lcdE; lit; int; xis; ycfK; ycfE; pin; ycfA.
SOURCE   Escherichia coli(strain:K12) DNA. clone:Kohara clone #240.
  ORGANISM Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 20284)
  AUTHORS   Mori,H.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
            Hirotada Mori, NARA Institute of Science and Technology, Res. &
            Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01,
            Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
REFERENCE   2 (sites)
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE   The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished (1996)
REFERENCE   3 (sites)
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE   A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map

```
          S19212"
          /codon_start=1
          /transl_table=11
          /product="Adenylosuccinate lyase (EC  4.3.2.2)"
          /protein_id="BAA35962.1"
          /db_xref="GI:1651565"
          /translation="MELSSLTAVSPVDGRVGDRVSALRGIFSExGLLKFRVQVEVRWL
          QKLAAHAIKEVPFAADAIGYLDAIVASFSEEDAARIKTIERTNHDVKAVEYFLKE
          KVAEIPELHAVSEFIHFACTSEDINNLSHALMLKTARDEVILPYWRQLIDGIKDLAVQ
          YRDIPLLSRTHGQPATPSTIGKEMANVAYRMERQYRQLNQVEILGKINGAVGNYNAHI
          AAYPEVDWHQFSEEFVTSLGIQWNPYTTQIEPHDYIAELFDCVARFNTLLIDFDRDVW
          GYIALNHFKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNAVLQHLASKLPVSRWQ
          RDLTDSTVLRNLGVGIGYRALIAYQSTLKGVSKLEVNRDHLLDELDHNWEVLAEPIQTV
          MRRYGIEKPYEKLKELTRGKRVDAEGMKQFIDGLALPEEEKARLKAMTPANYIGRAIT
          MVDELK"
gene      complement(2541..3182)
          /gene="ycfC"
CDS       complement(2541..3182)
          /gene="ycfC"
          /note="ORF_ID:o240#3; similar to PIR Accession Number
          S19211"
          /codon_start=1
          /transl_table=11
          /product="Hypothetical protein 23."
          /protein_id="BAA35963.1"
          /db_xref="GI:4062714"
          /translation="MAKNYDITLALAGICQSARLVQQLAHQGHCDADALHVSLNSII
          DMNPSTLAVFGGSEANLRVGLETLLGVLNASSRQGLNAELTRYTLSLMVLERKLSSA
          KGALDTLGNRINGLQRQLEHFDLQSETLMSAMAAIYVDVISPLGPRIQVTGSPAVLQS
          PQVQAKVRATLLAGIRAAVLWHQVGGRLQLMFSRNRLITQAKQILAHLTPEL"
gene      complement(3218..4324)
          /gene="ycfB"
CDS       complement(3218..4324)
          /gene="ycfB"
          /note="ORF_ID:o240#4; similar to SwissProt Accession
          Number P25745"
          /codon_start=1
          /transl_table=11
          /product="Hypothetical protein in purB 5'region (orf-15)
          ."
          /protein_id="BAA35964.1"
          /db_xref="GI:4062715"
          /translation="MSETAKKVIVGMSGGVDSSVSAWLLQQGYQVEGLFMKNWEEDD
          GEEYCTAAADLADAQAVCDKLGIELHTVNFAAEYWDNVFELFLAEYKAGRTPNPDILC
          NKEIKFKAFLEFAAEDLGADYIATGHYVRRADVDGKSRLLRGLDSNKDQSYFLYTLSH
          EQIAQSLFPVGELEKPQVRKIAEDLGLVTAKKDSTGICFIGERKFREFLGRYLPAQP
          GKIITVDGDEIGEHOQLMYHTLGQRKGLGIGGTKEGTEEPWYVVDKDVENNILVAQG
          HEHPRLMSVGLIAQQLHWDREPFTGTMRCTVKIRYRQTDIPCTVKALDDDRIEVIFD
          EPVAAVTPGQSAVFYNGEVCLGGGIIEQRLPLPV"
gene      complement(4378..4839)
CDS       complement(4378..4839)
          /note="ORF_ID:o240#5; similar to PIR Accession Number
          S64074"
          /codon_start=1
          /transl_table=11
          /product="Hypothetical protein YGL067w"
          /protein_id="BAA35965.1"
          /db_xref="GI:4062716"
          /translation="MFKPHYTVACVVHAEGKFLVVEETINGKALWNQPAGHLEADETL
          VEAAARELWEETGISAQPQHFIRMHQWIAPDKTPFLRFLFAIELEQICPTQPHDSDID
          CCRWVSAEEILQASNLRSPLVAESIRCYQSGQRYPLEMIGDFNWPFTKGVI"
gene      complement(4849..5502)
CDS       complement(4849..5502)
          /note="ORF_ID:o240#6; similar to PIR Accession Number
          I64156"
          /codon_start=1
          /transl_table=11
          /product="Hypothetical protein HI0694"
          /protein_id="BAA35966.1"
          /db_xref="GI:4062717"
          /translation="MRQFIISENTMQKTSFRNHQVKRFSSQRSTRRKPENQPTRVILF
          NKPYDVLPQFTDEAGRKTLKEFIPVGVYAAGRLDRDSEGLLVLTNNGALQARLTQPG
          KRTGKIYYVQVEGIPTQDALEALRNGVTLNDGPTLPAGAELVDEPAWLWPRNPPIRER
          KSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLJRYAMGDYSLDNLANGEWREVTD"
gene      5674..6924
```

Right column:

```
          /gene="icd"
          5674..6924
          /gene="icd"
CDS       /note="ORF_ID:o240#7; similar to SwissProt Accession
          Number P08200"
          /codon_start=1
          /transl_table=11
          /product="Isocitrate dehydrogenase (NADP) (EC  1.1.1.42)
          (oxalosuccinate decarboxylase) (idh) (NADP+-specific icdh)
          (idp)."
          /protein_id="BAA35967.1"
          /db_xref="GI:1651566"
          /translation="MESKVVPAQGKKITLQNGKLNVPENPIPYIEGDGIGVDVTPA
          MLKVVDAAVEKAYKGERKISWMEIYTGEKSTQVYGODVWLPAETLDLIREYRVAIKGP
          LTTPVGGGIRSLNVALRQELDLYICLRPVRYQGTPSPVKHPELTDMVIFRENSEDIY
          AGIEWKADSADAEKVIKFLREEMGVKKIRFPEHCGIGIKPCSEEGTKRLVRAAIEYAI
          ANDRDSVTLVHKGNIMKFTEGAFKDWGYQLAREEFGGELIDGGPWLKVKNPNTGKEIV

alignment_scores:
          Quality:     983.00          Length:       218
          Ratio:         4.795          Gaps:           0
Percent Similarity:    94.037    Percent Identity:  85.780

alignment_block:
US-09-252-691-7056 x D90749/rev

Align seg 1/1   to reverse of: D90749  from:  1   to: 20284

     1 AlaAlaMetArgGlnLeuIleThrProGluAsnThrMetThrLysThrSe    17
       |||||||||||||||::||||         ||||||||:||||||   ||||||
  5508 GCTATAATGGGGCAATTCATAATCTCGAAAATACCATGCAAAAACTTC  5459

    17 rPheArgLysHisArgValGluArgPheSerSerArgGlnAlaThrArgA    34
       ::||||||||:::|||:|:||||::||||||::|||||::||::||||||
  5458 TTTTAGAAATCACCAGGTTAAGGCATTCAGCTCGCCAACGTTCTACCAGGC  5409

    34 rgThrProGlnProGlnProArgValIleLeuPheAsnLysProTyr    50
       ||     |||||||||||||||||:||||||||||||||||||||||
  5408 GTAAACCTGAAAACCAGCCCACGCGTGTGATCCTGTTCAATAAACCCTAC  5359

    51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs    67
       |||||||||||||||||||||||||||||:::||::||:||||:|||::
  5358 GATGTTCTTCCGCAGTTCACCGATGAGGCCGGACGCAAAACATTAAAAGA  5309

    67 pPheIleProValGlnGlyValTyrAlaAlaLaGlyArgLeuAspArgAspS    84
       ::|||||||||||||:||::||  :::||||::||||||||::||::|||||
  5308 ATTCATCCCGGTTCAGGGTGTTTATGCAGCAGGTCGCCTTGACCCGCGATA  5259

    84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu   100
       ||||||||||||||||:|||::||||||:||||||||||||||||||||
  5258 GCGAAGGTTGCTGGTGCTGCTGACCAATAACGGCGCGTTGCAGGCGCGTTTA  5209

   101 ThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGl   117
       ||||||||||||||||||||||||||||||:||||||::||||||||||
  5208 ACCCAGCCGGGTAAACCCACCGGGAAAATCTATTATGTGCAGGTGGAAGG  5159

   117 yGluProAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuA   134
       |||::||   |::  :::||   |||||||||||||||:||||||||||||
  5158 TATTCCCACACAAGACGCACTTGAGCCTTGGCCAATGGGCAATGGGCTAACCTTAA  5109

   134 snAspGlyProThrLeuProAlaGlyAlaGluLeuValAspGluProGlu   150
       ||||||||||||||||||||||::|||||||||||  ::|::|||||
  5108 ATGATGGCCCTACCCTGCCCGCCGGGGCCGGAGCTGGTTGACGAACCCGCG  5059

   151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProTh   167
       ||||||||||||||||||||:|||::||||||||||||||||||||::|||
  5058 TGGTTATGGCCCGGAATCCACCAATTCGTGAACGCAAAAGTATTCCCAC  5009

   167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA   184
       ::|||||||||||||||||:||||:||||::||||||||||||||||||
  5008 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGCC  4959

   184 rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla   200
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  5008 ???
```

```
4958 GCATGACCGCCCATGTTGGCTTCCCCACGCTGCGACTGATTCGCTATGCG 4909
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 201 MetGlySerTyrThrLeuAspAsnAlaAsnGlyGluTrpArgArgAspVa 217
     ||||:::||||:::||||||::||||||||||||||||||||||||:::||
4908 ATGGGTGATTACTCTTTGGATAATCTTGCCAATGTGAATGGCCAGAAGT 4859

 217 1Thr 218
     ||||
4858 GACA 4855

seq_name: gb_htg7:AC022161

seq_documentation_block:
LOCUS       AC022161    180993 bp    DNA         HTG        26-JAN-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-273P11, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC022161
VERSION     AC022161.1  GI:6758614
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180993)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 180993)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
          * NOTE: This record contains 113 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.
          *        1  511: contig of 511 bp in length
          *      512 1507: gap of unknown length
          *     1508 2279: contig of 996 bp in length
          *     2280 2614: contig of 772 bp in length
          *     2615 3450: gap of 335 bp in length
          *     3451 3914: contig of 836 bp in length
          *     3915 4858: gap of unknown length
          *     4859 5380: contig of 464 bp in length
          *     5381 6247: gap of 944 bp in length
          *     6248 7203: contig of 522 bp in length
          *     7204 7920: gap of 867 bp in length
          *     7921 8924: contig of 956 bp in length
          *                gap of 717 bp in length
          *                contig of 1004 bp in length
          *                gap of unknown length
```
```
          *     8925  9674: contig of 750 bp in length
          *     9675 11049: gap of unknown length
          *    11050 11834: contig of 1375 bp in length
          *    11835 12774: contig of 785 bp in length
          *    12775 13555: gap of unknown length
          *    13556 14538: contig of 940 bp in length
          *    14539 15398: gap of unknown length
          *    15399 15696: contig of 781 bp in length
          *    15697 16675: gap of unknown length
          *    16676 17597: contig of 983 bp in length
          *    17598 18969: gap of unknown length
          *    18970 19300: contig of 860 bp in length
          *    19301 19380: gap of unknown length
          *    19381 20457: contig of 298 bp in length
          *    20458 21187: contig of 979 bp in length
          *    21188 21435: gap of unknown length
          *    21436 22621: contig of 922 bp in length
          *    22622 23209: gap of unknown length
          *    23210 23994: contig of 1372 bp in length
          *    23995 24643: gap of unknown length
          *    24644 25421: contig of 331 bp in length
          *    25422 25826: gap of unknown length
          *    25827 27532: contig of 80 bp in length
          *    27533 28937: gap of unknown length
          *    28938 30011: contig of 1077 bp in length
          *    30012 30574: gap of unknown length
          *    30575 31545: contig of 730 bp in length
          *    31546 32314: gap of unknown length
          *    32315 33010: contig of 248 bp in length
          *    33011 33654: gap of unknown length
          *    33655 34705: contig of 1186 bp in length
          *    34706 35755: gap of unknown length
          *    35756 36672: contig of 588 bp in length
          *    36673 37904: gap of unknown length
          *    37905 38457: contig of 785 bp in length
          *    38458 39301: gap of unknown length
          *    39302 40257: contig of 649 bp in length
          *                 contig of 778 bp in length
          *                 gap of unknown length
          *                 contig of 405 bp in length
          *                 gap of unknown length
          *                 contig of 1706 bp in length
          *                 gap of unknown length
          *                 contig of 1405 bp in length
          *                 gap of unknown length
          *                 contig of 1074 bp in length
          *                 gap of unknown length
          *                 contig of 563 bp in length
          *                 gap of unknown length
          *                 contig of 971 bp in length
          *                 gap of unknown length
          *                 contig of 769 bp in length
          *                 gap of unknown length
          *                 contig of 696 bp in length
          *                 gap of unknown length
          *                 contig of 644 bp in length
          *                 gap of unknown length
          *                 contig of 1051 bp in length
          *                 gap of unknown length
          *                 contig of 1050 bp in length
          *                 gap of unknown length
          *                 contig of 917 bp in length
          *                 gap of unknown length
          *                 contig of 1232 bp in length
          *                 gap of unknown length
          *                 contig of 553 bp in length
          *                 gap of unknown length
          *                 contig of 844 bp in length
          *                 gap of unknown length
          *                 contig of 956 bp in length
```

```
*  40258    gap of unknown length      41691:  contig of 1434 bp in length
*  41692    gap of unknown length      43227:  contig of 1536 bp in length
*  43228    gap of unknown length      44416:  contig of 1189 bp in length
*  44417    gap of unknown length      45206:  contig of 790 bp in length
*  45207    gap of unknown length      46223:  contig of 1017 bp in length
*  46224    gap of unknown length      47104:  contig of 881 bp in length
*  47105    gap of unknown length      49150:  contig of 2046 bp in length
*  49151    gap of unknown length      50227:  contig of 1077 bp in length
*  50228    gap of unknown length      51990:  contig of 1763 bp in length
*  51991    gap of unknown length      52651:  contig of 661 bp in length
*  52652    gap of unknown length      54157:  contig of 1506 bp in length
*  54158    gap of unknown length      55684:  contig of 1527 bp in length
*  55685    gap of unknown length      56936:  contig of 1252 bp in length
*  56937    gap of unknown length      57993:  contig of 1057 bp in length
*  57994    gap of unknown length      58875:  contig of 882 bp in length
*  58876    gap of unknown length      60147:  contig of 1272 bp in length
*  60148    gap of unknown length      61314:  contig of 1167 bp in length
*  61315    gap of unknown length      61502:  contig of 188 bp in length
*  61503    gap of unknown length      62298:  contig of 796 bp in length
*  62299    gap of unknown length      63529:  contig of 1231 bp in length
*  63530    gap of unknown length      65376:  contig of 1847 bp in length
*  65377    gap of unknown length      65461:  contig of 85 bp in length
*  65462    gap of unknown length      66798:  contig of 1337 bp in length
*  66799    gap of unknown length      68065:  contig of 1267 bp in length
*  68066    gap of unknown length      69742:  contig of 1677 bp in length
*  69743    gap of unknown length      71264:  contig of 1522 bp in length
*  71265    gap of unknown length      72180:  contig of 916 bp in length
*  72181    gap of unknown length      74061:  contig of 1881 bp in length
*  74062    gap of unknown length      75047:  contig of 986 bp in length
*  75048    gap of unknown length      76521:  contig of 1474 bp in length
*  76522    gap of unknown length      78628:  contig of 2107 bp in length
*  78629    gap of unknown length      80347:  contig of 1719 bp in length
*  80348    gap of unknown length      82199:  contig of 1852 bp in length
*  82200    gap of unknown length      83804:  contig of 1605 bp in length
```

alignment_scores:
Quality: 850.00   Length: 193
Ratio: 4.749   Gaps: 1
Percent Similarity: 92.746 Percent Identity: 85.492

alignment_block:
US-09-252-691-7056 x AC022161

Align seg 1/1 to: AC022161 from: 1 to: 180993

```
1    AlaIleMetArgGlnLeuIleThrProGluAsnThrMetThrLysThrSe    17
     |||||||||||||| ||||||||||||| |:||||||::|||||||
31737 GCTATAATGCGGCAATTCATAATCTCTGAAAATACCATGCAAAAAACTTC   31786

17   rPheArgLysHisArgValGluArgPheSerArgGlnAlaThrArgA    34
     ||||:|||:|||::||::|||||||:||||||:::|||::||:|||||||
31787 TTTTAGAAATACCAGGTTAAGCGATTCAGCTCGCCAACGTTCTACCAGGC   31836

34   rgThrProGlnProThrArgValIleLeuPheAsnLysProTyr    50
     ||||||||||||::||||||||||||||||||||||
31837 GTCAACCTGAAAACCAGCCCACGCGTGATCCTGTTCAATAAACCCTAC    31886

51   AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs    67
     |||||||||||||||||||||||::|||||:||||||::::||||||
31887 GATGTTCTTCCGCAGTTCACCGATGAAGCCGGACCGCAAAACATTAAAAGA   31936

67   pPheIleProValGlnGlyValTyrAlaAlaAlaGlyArgLeuAspArgAspS    84
     :||||||||||||||||||||| |||||||:::|||||||:::|||
31937 ATTCATCCCGGTTCAGGGTGTTATGCAGCAGGTCGCCTTGACCGCCGATA   31986

84   erGluGlyLeuLeuValLeuThrArgAsnAspGlyValLeuGlnAlaArgLeu    100
     ||||||||||||||||||||||||||:|||||::|||||||||||
31987 GCGAAGGGTTGCTGGTCCTGACCAATAACGGCGCGTTGCAGGCGCGTTTA   32036

101  ThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGl    117
     |||||||||||||||||||||||||||| ||||||||||||||
32037 ACCCAGCCGGGTAAACGCACCGGAAAAATCTATTATGTGCAGGTGGAAGG   32086

117  y.GluProAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeu    133
      |||  ::||   ||||||||||||||||||||||||||
32087 TATTCCCCACAAGACGCACTGAAGCCTTGCGCAATGGCGTAACCTTA    32136

134  AsnAspGlyProThrLeuProAlaGlyIleIleGluArgValAsnGluProGl    150
     ||||  ||||  ||||||||  ||  ||:|||||||||||||||
32137 AATGAATGGCCCTACCCTGCCCGCCGGCGGAGCTGGTTGACGAACCCGC   32186

150  uTrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProT    167
     |||||||||||||||||||||||||||||||||||||||||||
32187 GTGGTTATGGCCCGCGGAATCCACCAATTCGTGAACGCAAAAGTATTCCCA   32236

167  hrSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArg    183
     ||||||||||||||||||||||||||||||||||||||||||||
32237 CCAGCTGGCTGGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGC   32286

184  ArgMetThrAlaHisValGlyPhePro    192
     |||||||||||||||||||||||||
32287 CGCATGACCGCCCATGTTGGCTTCCCC    32313
```

seq_name: gb_htg7:AC009127

TITLE
JOURNAL   Direct Submission
          Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Jan 26, 2000 this sequence version replaced gi:5685948.
          ------Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          ----
          * NOTE: This record contains 101 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.

```
*          1       989: contig of 989 bp in length
*        990      1318: gap of 329 bp in length
*       1319      1843: contig of 525 bp in length
*       1844      2066: gap of unknown length
*       2067      3127: contig of 1061 bp in length
*       3128      3535: gap of unknown length
*       3536      4171: contig of 636 bp in length
*       4172      4797: gap of unknown length
*       4798      5170: contig of 373 bp in length
*       5171      5390: gap of unknown length
*       5391      6021: contig of 631 bp in length
*       6022      6686: gap of unknown length
*       6687      6975: contig of 289 bp in length
*       6976      7331: gap of unknown length
*       7332      7911: contig of 580 bp in length
*       7912      8248: gap of unknown length
*       8249      8844: contig of 596 bp in length
*       8845      9149: gap of unknown length
*       9150      9469: contig of 320 bp in length
*       9470      9911: gap of unknown length
*       9912     10494: contig of 583 bp in length
*      10495     11143: gap of unknown length
*      11144     11762: contig of 619 bp in length
*      11763     12380: gap of unknown length
*      12381     13000: contig of 620 bp in length
*      13001     13576: gap of unknown length
*      13577     14175: contig of 599 bp in length
*      14176     14801: gap of 626 bp in length
*      14802     15373: contig of 572 bp in length
*      15374     15716: gap of unknown length
*      15717     16489: contig of 773 bp in length
*      16490     17111: gap of unknown length
*      17112     17855: contig of 744 bp in length
*      17856     19178: gap of unknown length
*      19179     20052: contig of 874 bp in length
*      20053     20790: gap of unknown length
*      20791     21464: contig of 674 bp in length
*      21465     22165: gap of unknown length
*      22166     22966: contig of 801 bp in length
*      22967     23458: gap of unknown length
*      23459     24138: contig of 680 bp in length
*      24139     24984: gap of unknown length
*      24985     25642: contig of 658 bp in length
*      25643     26565: gap of unknown length
*      26566     27196: contig of 631 bp in length
*      27197     27815: gap of unknown length
*      27816     27901: contig of 86 bp in length
*      27902     28512: gap of unknown length
*      28513     29881: contig of 1369 bp in length
*      29882     30600: gap of unknown length
*      30601     31204: contig of 604 bp in length
*      31205     32333: gap of unknown length
*      32334     32911: contig of 578 bp in length
*      32912     33816: gap of unknown length
*      33817     34778: contig of 962 bp in length
*      34779     36264: gap of unknown length
*      36265     38613: contig of 2349 bp in length
*      38614     39144: gap of unknown length
*      39145     41224: contig of 2080 bp in length
*      41225     41982: gap of unknown length
*      41983     43378: contig of 1396 bp in length
*      43379     44856: gap of unknown length
*      44857     46513: contig of 1657 bp in length
```

```
  101 ThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGl 117
      |||||||||||||||||||||||||||||||||||||||||||||||||||
28248 ACCCAGCCGGTAAACGCACCGGAAAAATCTATTATGTGCAGGTGGAAGG 28297

  117 yGluProAspAspAlaSerLeuAlaLysArgAsnGlyValThrLeuA 134
      ||| :::|||   ::|||        ||||||||||||||||||||||
28298 TATTCCACACAAGACGCACTTGAAGCCTTGCGCAATGGCGTAACCTTAA 28347

  134 snAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProGlu 150
      |||||||||||||||||||||||||||   |||::::||||||||||||
28348 ATGATGGCCCTACCCCGCCCGGCGGGAGCTGTTGACGAACCGCG 28397

  151 TrpLeuTrpProArgAsnProProIleArgGluArgGlySerIleProTh 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
28398 TGGTTATGGCGCGGAATCACCAATTCGTGAACGACAAAGTATTCCCAC 28447

  167 rSerTrpLeuIleThrLeuTyrGluGlyValArgAsnArgGlnValArgA 184
      |||||||||||||||||||||||||||||||    ||||||||:::|||
28448 CAGCTGGCTTAAGATCACCTTATATGAAAGACGTAATCGGCAGTTGCGCC 28497

  184 rgMetThrAlaHis 188
      ||||||||||||||
28498 GCATGACGGCCCAT 28511

seq_name: gb_htg6:AC008911

seq_documentation_block:
LOCUS        AC008911    226805 bp    DNA                    HTG        14-JAN-2000
DEFINITION   Homo sapiens chromosome 5 clone CITB-H1_2228J5, WORKING DRAFT
             SEQUENCE, 95 unordered pieces.
ACCESSION    AC008911
VERSION      AC008911.2    GI:6693307
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 226805)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Sequencing of Human Chromosome 5
  JOURNAL    Unpublished
REFERENCE    2 (bases 1 to 226805)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Direct Submission
  JOURNAL    Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      On Jan 14, 2000 this sequence version replaced gi:5686162.
             ------Genome Center
             Center: Joint Genome Institute
             Center Code: JGI
             Web site: http://www.jgi.doe.gov
             ------Summary Statistics
             Consensus quality: 120029 bases at least Q40
             Consensus quality: 163743 bases at least Q30
             Consensus quality: 177506 bases at least Q20
             Estimated insert size: 226805; sum-of-contigs estimation
             Estimated insert size: 164170; agarose-fp estimation
             Quality coverage:  4.03x in Q20 bases; agarose-fp estimation
             Quality coverage:  2.91x in Q20 bases; sum-of-contigs estimation
             ------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 95 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1    1018: contig of 1018 bp in length
             *     1019    2070: gap of unknown length
             *     2071    3108: contig of 1052 bp in length
             *                   gap of unknown length
             *                   contig of 1038 bp in length
```

```
     *   46514  48278: contig of 1765 bp in length
     *          48279: gap of unknown length
     *   49123: contig of 845 bp in length
     *          gap of unknown length
     *   49124  50185: contig of 1062 bp in length
     *   50186          gap of unknown length
     *   51798: contig of 1613 bp in length
     *          gap of unknown length
     *   51799  52794: contig of 996 bp in length
     *   52795          gap of unknown length
     *   54154: contig of 1360 bp in length
     *          gap of unknown length
     *   54155  56271: contig of 2217 bp in length
     *   56272          gap of unknown length
     *   58993: contig of 2722 bp in length
     *          gap of unknown length
     *   58994  60268: contig of 1275 bp in length
     *   60269          gap of unknown length
     *   62633: contig of 2365 bp in length
     *          gap of unknown length
     *   62634  65099: contig of 2466 bp in length
     *   65100          gap of unknown length
     *   66776: contig of 1677 bp in length
     *          gap of unknown length
     *   66777  68874: contig of 2098 bp in length
     *          gap of unknown length
     *   68875  72266: contig of 3392 bp in length
     *   72267          gap of unknown length
     *   74691: contig of 2425 bp in length
     *          gap of unknown length
     *   74692  77588: contig of 2897 bp in length
     *   77589          gap of unknown length
     *   80107: contig of 2519 bp in length
     *          gap of unknown length
     *   80108  84825: contig of 4718 bp in length
     *   84826          gap of unknown length
     *   87029: contig of 2204 bp in length
     *          gap of unknown length

alignment_scores:
     Quality:          824.00         Length:          188
     Ratio:            4.763          Gaps:            0
Percent Similarity: 92.021         Percent Identity: 84.043

alignment_block:
US-09-252-691-7056 x AC009127
Align seg 1/1   to: AC009127   from: 1   to: 185591

    1 AlaIleMetArgGlnLeuIleThrProGluAsnThrMetThrLysThrSe 17
      ||||||||||||||:::|||        ||||||  |||||
27948 GCTATAAATGCGGCAATTCATAATCTCTGAAAATACCATGCAAAAAACTTG 27997

   17 rPheArgLysHisArgValGluArgPheSerSerArgGlnAlaThrArgA 34
      |||||||||||||::||:::|:||::||||||||||||||||||:::|||
27998 TTTTAGAAATCACCAGGTTAAGCAATTCAGCTCGCAACGTTCTACCAGGC 28047

   34 rgThrProGluProGlnProThrArgValIleLeuPheAsnLysProTyr 50
      || |||||  |||||||||||||||||||||||||||||||||||
28048 GTAAACCTGAAACCAGCCCACGCGTGTGATCCTGTTCAATAAACCCTAC 28097

   51 AspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAs 67
      ||||||||||||||||||||||||:::|:||:|||||||||||::
28098 GATGTTCTTCCGCAGTTCACCGATGAAGCCGGACGCAAAACATTAAAGA 28147

   67 pPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspS 84
      :||||||||||||||||||||||||||||||||||||||||||||||||
28148 ATTCATCCGGTTCAGGGTGTTTATGCAGCAGGTCGCCTTGACCGCGATA 28197

   84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
      |||||||||||||||||||||||:::|:||||||||||||||||||||||
28198 GCGAAGGGTTGCTGGTGCTGACCAATAACGGCGCCGTTGCAGGCGCGTTTA 28247
```

us-09-252-691-7056.rge

Tue Jun 13 10:30:08 2000

```
   3109 *  gap of unknown length
   4211:   contig of 1103 bp in length
   4212 *  gap of unknown length
   5226:   contig of 1015 bp in length
   5227 *  gap of unknown length
   6336:   contig of 1110 bp in length
   6337 *  gap of unknown length
   7420:   contig of 1084 bp in length
   7421 *  gap of unknown length
   8436:   contig of 1016 bp in length
   8437 *  gap of unknown length
   9514:   contig of 1078 bp in length
   9515 *  gap of unknown length
  11442:   contig of 1928 bp in length
  11443 *  gap of unknown length
  12908:   contig of 1466 bp in length
  12909 *  gap of unknown length
  14243:   contig of 1335 bp in length
  14244 *  gap of unknown length
  15370:   contig of 1127 bp in length
  15371 *  gap of unknown length
  16374:   contig of 1004 bp in length
  16375 *  gap of unknown length
  17484:   contig of 1110 bp in length
  17485 *  gap of unknown length
  18486:   contig of 1002 bp in length
  18487 *  gap of unknown length
  19535:   contig of 1049 bp in length
  19536 *  gap of unknown length
  20747:   contig of 1212 bp in length
  20748 *  gap of unknown length
  21794:   contig of 1047 bp in length
  21795 *  gap of unknown length
  22910:   contig of 1116 bp in length
  22911 *  gap of unknown length
  24510:   contig of 1600 bp in length
  24511 *  gap of unknown length
  26330:   contig of 1820 bp in length
  26331 *  gap of unknown length
  27876:   contig of 1546 bp in length
  27877 *  gap of unknown length
  29050:   contig of 1174 bp in length
  29051 *  gap of unknown length
  30329:   contig of 1279 bp in length
  30330 *  gap of unknown length
  31614:   contig of 1285 bp in length
  31615 *  gap of unknown length
  33078:   contig of 1464 bp in length
  33079 *  gap of unknown length
  34288:   contig of 1210 bp in length
  34289 *  gap of unknown length
  35341:   contig of 1053 bp in length
  35342 *  gap of unknown length
  36695:   contig of 1354 bp in length
  36696 *  gap of unknown length
  37729:   contig of 1034 bp in length
  37730 *  gap of unknown length
  39224:   contig of 1495 bp in length
  39225 *  gap of unknown length
  40393:   contig of 1169 bp in length
  40394 *  gap of unknown length
  41828:   contig of 1435 bp in length
  41829 *  gap of unknown length
  43639:   contig of 1811 bp in length
  43640 *  gap of unknown length
  45023:   contig of 1384 bp in length
  45024 *  gap of unknown length
  46157:   contig of 1134 bp in length
  46158 *  gap of unknown length
  47250:   contig of 1093 bp in length
  47251 *  gap of unknown length
  48869:   contig of 1619 bp in length

  48870 *  gap of unknown length
  50138:   contig of 1269 bp in length
  50139 *  gap of unknown length
  51169:   contig of 1031 bp in length
  51170 *  gap of unknown length
  52255:   contig of 1086 bp in length
  52256 *  gap of unknown length
  53454:   contig of 1199 bp in length
  53455 *  gap of unknown length
  55052:   contig of 1598 bp in length
  55053 *  gap of unknown length
  56384:   contig of 1332 bp in length
  56385 *  gap of unknown length
  57426:   contig of 1042 bp in length
  57427 *  gap of unknown length
  58894:   contig of 1468 bp in length
  58895 *  gap of unknown length
  59899:   contig of 1005 bp in length
  59900 *  gap of unknown length
  61097:   contig of 1198 bp in length
  61098 *  gap of unknown length
  62827:   contig of 1730 bp in length
  62828 *  gap of unknown length
  64143:   contig of 1316 bp in length
  64144 *  gap of unknown length
  65320:   contig of 1177 bp in length
  65321 *  gap of unknown length
  66338:   contig of 1018 bp in length
  66339 *  gap of unknown length
  67938:   contig of 1600 bp in length
  67939 *  gap of unknown length
  69534:   contig of 1596 bp in length
  69535 *  gap of unknown length
  70638:   contig of 1104 bp in length
  70639 *  gap of unknown length
  71721:   contig of 1083 bp in length
  71722 *  gap of unknown length
  72737:   contig of 1016 bp in length
  72738 *  gap of unknown length
  74240:   contig of 1503 bp in length
  74241 *  gap of unknown length
  76005:   contig of 1765 bp in length
  76006 *  gap of unknown length
  77033:   contig of 1028 bp in length
  77034 *  gap of unknown length
  79330:   contig of 2297 bp in length
  79331 *  gap of unknown length
  80402:   contig of 1072 bp in length
  80403 *  gap of unknown length
  82066:   contig of 1664 bp in length
  82067 *  gap of unknown length
  83385:   contig of 1319 bp in length
  83386 *  gap of unknown length
  85506:   contig of 2121 bp in length
  85507 *  gap of unknown length
  87079:   contig of 1573 bp in length
  87080 *  gap of unknown length
  88558:   contig of 1479 bp in length
  88559 *  gap of unknown length
  89799:   contig of 1241 bp in length
  89800 *  gap of unknown length
  92035:   contig of 2236 bp in length
  92036 *  gap of unknown length
  94093:   contig of 2058 bp in length
  94094 *  gap of unknown length
  96588:   contig of 2495 bp in length
  96589 *  gap of unknown length
  98564:   contig of 1976 bp in length
  98565 *  gap of unknown length
 100900:   contig of 2336 bp in length
 100901 *  gap of unknown length
 103171:   contig of 2271 bp in length
 103172 *  gap of unknown length
 105268:   contig of 2097 bp in length
```

```
alignment_scores:
*       105269    108853: contig of 3585 bp in length
                          gap of unknown length
*       108854    112992: contig of 4139 bp in length
                          gap of unknown length
*       112993    117661: contig of 4669 bp in length
                          gap of unknown length
*       117662    121578: contig of 3917 bp in length

        Quality:       696.50        Length:       234
        Ratio:           4.003       Gaps:           4
Percent Similarity:    74.359        Percent Identity:   62.821

alignment_block:
US-09-252-691-7056 x AC008911/rev

Align seg 1/1  to reverse of: AC008911  from: 1  to: 226805
```

```
    10 GluAsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPh 26
       ::|||||||||||||||||||||||||||||:|||::::|||:|||||||
 51170 GAAAATACCATGCAAAAAACTTCTTTAGAAAATCACCAGGTTAAGCGATT 51121

    26 eSerSerArgGlnAlaThrArgArgThrProGlnProThrArgV 43
       |||||:::|||::||||||||||||| |||||| ||||||||||
 51120 CAGCTGCCAACGTTCTACCAGGGGCTAAACCTGAAAACCACGCGTG 51071

    43 aIleLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGlu 59
       |||||||||||||||||||||||||||||||||||||||||||||||||
 51070 TGATCCTGTTCAATAAACCCTACGATGTTCTCCGCAGTTCACCGATGAA 51021

    60 AlaGlyArgSerThrLeuLysAspPheIleProValGlnGlyValTyrA1 76
       ||||||||||||||||||||:||||||||||||||||||||||||||||
 51020 GCCGGACGCAAAACATTAAAGAATTCATCCCGGTTCAGGGTGTTTATGC 50971

    76 aAlaGlyArgLeuAspArgAspSerGluLeuValLeuThrAsnA 93
       |||||||||||||||||||||||||||||||||||||||||||||
 50970 AGCAGGTCGCCTTGACCGGGATAGCGAAGGGTTCGTGGTCTGACCAATA 50921

    93 spGlyValLeuGlnAlaArgLeuThr...GlnProGlyLysArgThrGly 108
       :::|||||||||||  ||||||||||      |||||:  :::||||
 50920 ACGCGGCGTTGCAGGCAGCGTTAACCCNNAGCCGGGNNTAAACGCACCG 50871

   109 Lys......IleTyrTyrValGlnValGlnGlyGluProAspAspAlaSe 123
       :::      ||||||||||||   |||  :: ||||||||||||
 50870 GAAANNNACTATTANTGTGCNAGGTGGAAGGTATTCCCACACAGAGCGC 50821

   123 rLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspAspProThrLeuP 140
       |||||||||||||||||||||||:|:||||||||||||||||||||||||
 50820 ACTTGAAGCCTTGCGCCAATGCGTAACCTTAAATGATGGCCCTACCCTGC 50771

   140 roAlaGlyIleGluArgValAlaAsnGluProGluTrpProProArgAsn 156
       |||||||  ||| |::|:|||||||||||||||||||||||||||||||
 50770 CCGCCGGCGCGGGAGCTGGTTGACGAACCCGCGTGTTATGGCCGCGGAAT 50721

   157 ProProIleArgGluArgLysSerIleProThrSerTrpLeuLysIleTh 173
       |||||||||||||||||||||||||||||||||||||||||||||||||
 50720 CCACCAATTCGTGAACGCAAAAGTATTCCCACAGCTGGCTGAAGATCAC 50671

   173 rLeuTyrGluGlyArg..............AsnArgGlnValA 183
       |||||||||||||||                  :::||||||||:
 50670 CTTATATGAAGGACGCCTAAGTCTCCAAGGGTTCGCACACGCAAGTGG 50621

   183 rgArgMetThrAlaHisValGlyPheProThrArgLeuIleArgTyr 199
       ||:||   :::    |||||:
 50620 AGCGCGGTGTCCGCATTGGCNTNGGGCCCTGTGCGCCACGCAAGCGGCCGTG 50571

   200 Ala......................Me 201
       |||                        ::
 50570 GCCGGCAGCGTGGACTCTNCGGCCTAAGTAGTGNCGAACTGGCGCTGACTT 50521

   201 tGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAspProValT 218
       :::||::::          |||  ||||||||||:::|||||
```

```
50520 GAGCTCGTTTGGAAATAAATGCTTCGCCAATGGTCGAATGCGAGAAGTGA 50471
                                                      ||
   218 hr 218
                                                      ||
50470 CA 50469

seq_name: gb_ba2:U32752

seq_documentation_block: 11833 bp   DNA
LOCUS        U32752        11833 bp    DNA            BCT    29-MAY-1998
DEFINITION   Haemophilus influenzae Rd section 67 of 163 of the complete genome.
ACCESSION    U32752 L42023
VERSION      U32752.1  GI:1573692
KEYWORDS     .
SOURCE       Haemophilus influenzae Rd.
  ORGANISM   Haemophilus influenzae Rd
             Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
             Haemophilus.
REFERENCE    1  (bases 1 to 11833)
  AUTHORS    Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
             Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
             Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
             Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
             Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
             Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
             Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
             Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
             Smith,H.O. and Venter,J.C.
  TITLE      Whole-genome random sequencing and assembly of Haemophilus
             influenzae Rd
  JOURNAL    Science 269 (5223), 496-512 (1995)
  MEDLINE    95350630
REFERENCE    2  (bases 1 to 11833)
  AUTHORS    Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
             Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE      Metabolism and evolution of Haemophilus influenzae deduced from a
             whole-genome comparison with Escherichia coli
  JOURNAL    Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE    96398784
REFERENCE    3  (bases 1 to 11833)
  AUTHORS    White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
REFERENCE    4  (bases 1 to 11833)
  AUTHORS    White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
  REMARK     The H. influenzae sequence has been updated by R Fleischmann. New
             database matches have been assigned, product names have been
             improved, and a number of frame shifts have been corrected. We
             gratefully acknowledge the work of Tatusov et. al. We have
             incorporated their annotation into the /notes fields of the
             corresponding H. influenzae genes
REFERENCE    5  (bases 1 to 11833)
  AUTHORS    White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
             Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
  REMARK     The whole genome was shifted by 588 nucleotides for a new start
COMMENT      On Sep 30, 1996 this sequence version replaced gi:1221393.
FEATURES             Location/Qualifiers
     source          1..11833
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            145..939
                     /gene="HI0690"
     CDS             145..939
                     /gene="HI0690"
                     /note="similar to GB:L19201 SP:P11244 GB:M55990 GB:U13915
                     GB:X15054 percent identity: 73.18; identified by sequence
```

```
                    similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="glycerol uptake facilitator protein (glpF)"
                    /protein_id="AAC22350.1"
                    /db_xref="GI:1573694"
                    /translation="MDKSLKANCIGEFLGTALLIFFGVGCVAALKVAGASFGLWEISI
                    MWGMGVALAVYATAGLSGAHLNPAVTIALWKFACFDGKKVIPYIISQMLGAFFAAALV
                    YALYRNVFIDYETVHNIVRGTQESLSLAGTFSTYPHPSLSIGGAFAVEFVTAILMAL
                    IMALTDDGNGVPRGPLAPLLIGILIAVIGGAMGPLTGFAMMPARDFGPKFFAYLAGWG
                    ELALTGGREIPYFIVPMVAPVLGALAGAWLIKKAICGNLPCNCGCE"
gene               960..2471
                   /gene="HI0691"
CDS                960..2471
                   /gene="HI0691"
                   /note="similar to GB:L19201 SP:P08859 GB:M18393 GB:M55990
                   GB:X15054 percent identity: 76.89; identified by sequence
                   similarity; putative"
                   /codon_start=1
                   /transl_table=11
                   /product="glycerol kinase (glpK)"
                   /protein_id="AAC22351.1"
                   /db_xref="GI:1573695"
                   /translation="MTDKKYIIALDQGTTSSRAVLLDHNANVVEIAQREFTQIYPRAG
                   WVEHNPMEIWATQSSTLNEVVMAKAGITSDEIAAIGITNQRETTIVWEKSTGTPVNAI
                   VWQCRRTADITDKLKADGHEEYIRNTTGLVVDPYFSGTKVKWILDNVEGAREKAERGE
                   LLFGTVDTWLVWKLTQGRVHVTDYTNASRTMLFNIHTKQWDDKMLEILNIPRSMLPEV
                   RNSSEIYGQTNIGGKGGVRIPVAGIAGDQQAALYGHLCVHAGQAKNTYGTGCFMLHT
                   GNKAITSKNGLLJTIACNAKGEPEYALEGSVFIAGASIQWLRDELKIVHDSFDSEYFA
                   QKVTDSNGVYVYVPAFTGLGAPYWDPYARGAIFGLSRGANRNHIVRATLESIAYQTRDV
                   LEAMQSDSGERLQYLRVDGGATNNFLMQFQADILDVNVERPVVKEVTALGAAYLAGL
                   ATGFWKDLDELRDKARVERTFSPDSDNEKRERRYKGWKKAVKRSLEWAKEDDE"
gene               complement(2546..3013)
                   /gene="HI0692"
CDS                complement(2546..3013)
                   /gene="HI0692"
                   /note="similar to SP:P26972 GB:X52093 GB:X63336 PID:47695
                   percent identity: 74.03; identified by sequence
                   similarity; putative"
                   /codon_start=1
                   /transl_table=11
                   /product="xanthine-guanine phosphoribosyltransferase
                   (gptB)"
                   /protein_id="AAC22352.1"
                   /db_xref="GI:1573693"
                   /translation="MSEKYVTWDMFQMHARRLSERLLPASQWKGIIAVSRGGLFPAA
                   VLARELGLRHIETVCIASYHDRNNQGELQVLHAAQVPNGGEGFIVVDDLVDTGNTARA
                   IRQMYPNAKFVTVFAKPAGAELVDDYVIDIPQNTWIEQPWDLGLIFVPPLSRK"
gene               3297..4121
                   /gene="HI0693"
CDS                3297..4121
                   /gene="HI0693"
                   /note="similar to GB:M68502 SP:P26093 PID:1573696 percent
                   identity: 99.64; identified by sequence similarity;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="lipoprotein E (hel)"
                   /protein_id="AAC22353.1"
                   /db_xref="GI:1573696"
                   /translation="MKTTLKMTALALSAFVLAGCGSHQMKSEGHANMQLQQQAVLGL
                   NWMQDSGEYKALAYQAYNAAKVAFDHAKVANGKKKAVVADLDETMLDNSPYAGWQVQN
                   NKPFDGKDWTRKVDARQSRAVPGAVEFENNYYNSHNGKVFYVTNRKDSTEKSGTIDDMK
                   RLGFNGVEESAFYLKKDKSAKAARFAEIEKOGYEIVLYVGDNLDDFGNTYVGLKNADR
                   RAFVDQNQGKFGKTFIMLPNANYGGWEGGLAEGYFKKDTQGOIKARLDAVQAWDGK"
gene               4261..4983
                   /gene="HI0694"
CDS                4261..4983
                   /gene="HI0694"
                   /note="similar to GB:U00096 SP:P75966 PID:1787380 percent
                   identity: 66.29; identified by sequence similarity;
                   putative"
                   /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAC22354.1"
                    /db_xref="GI:1573697"
                    /translation="MMLQAFQNRNHKLMKKNFSAGKLPSKGKSAVNFHRTFKPKLKPK
                    TLSLDETKVVLFNKPFDVLTQFTDEQGRATLKDFISIPNVYAAGRLDRDSEGLLILTN
                    NGELQHRLADPKFKTEKTYWVVGEIPEETDLAQLRKGVELKDGVTKSAKVRLISEPN
                    LWERNPIRERKNIPTSWLEIKISEGRNRQVRMTAHIGFPTLRIVRVSMGLLSINGL
                    ENGSFRLSLDEIKALFQTVKL"
gene               complement(5109..6080)
                   /gene="HI0695"
CDS                complement(5109..6080)
                   /gene="HI0695"
                   /note="similar to GB:L06129 SP:P29014 PID:147343 GB:U00096
                   PID:1788848 percent identity: 53.62; identified by
                   sequence similarity; putative"
                   /codon_start=1
                   /transl_table=11
                   /product="exopolyphosphatase, putative"
                   /protein_id="AAC22355.1"
                   /db_xref="GI:1573698"
                   /translation="MNDSILEPKHRGNVREIAIDLGSNSFHMVARIVNGSIQVLSR
                   LKQKVKLAEGLDENAVLNQEAITRGVNCLALFAERLQGFPMENVNVVGTYTLRRAVNN
                   DEFLRQAAKVFPYPINIISGQTEAKTIYAGVCHTOPEKGRKLVIDIGGGSTEMIIGDD
                   FTPLMAESRHMGCVSEIATGFTDGIISPENKAQSAESSRFFEDLGLEYRKLGWQSVL
                   GSSGTIKTVAQVIATNLDPNGTTTAERLNALIEQTLQAKHFTELNINGNQDRVDVFV
                   PGLAILSAVFDVFHIQMRYSDGALREGVIYSLEKNFQVADIRASTA"
gene               complement(6088..9984)
                   /gene="HI0696"
CDS                complement(6088..9984)
                   /gene="HI0696"
                   /note="similar to GB:U14003 SP:P39322 PID:537063 GB:U00096
                   percent identity: 34.87; identified by sequence
                   similarity; putative"
                   /codon_start=1
                   /transl_table=11
                   /product="conserved hypothetical protein"
                   /protein_id="AAC22356.1"
                   /db_xref="GI:1573699"
                   /translation="MTEQIQPSETSPKSPEKPNKKHWVRKAVCIGSAVILVPVLGVAG
                   ALSFDAGQKSLIQLVDKMLDSFSVEQVEGGLQNGLVLKNVRYQTAGIETHIAQARLQL
                   DFGCLFSREVCLRDFTLNKPTIAINTALLPPSAPDNSKSGSMKRISLPISINAENLVM
                   QDLSVNIDQTSITLGNFKSAVSLNNEKGLTIAPTEINDISVIAKKLSEVKSEPKAEQP
                   NKPVDWAAIEQSLTPAFLGNVSEIILPDLHIPEISGKNWQYQAVNEKGETLQSVEMS
                   SLIAQADTVDNQLQLQKLAVESSLGNLSSQGKLQLDGDMPLDLTLKSHLEPLKSDGKE
                   ILPASDVDLTLSGSLKKSTALSLKTKGVLDAELNGNVQLAQDKMPLNLTLNVAKGQYT
                   FVNTMTPLKINDVTLKLTGDLLNYHAELKGVAGMNYIPASQVELNADGKLYEVTVNK
                   LGIDSLDGKSEFVGNANWKNGANWDIQADLEKMNIAFFVPVMPATLSGKLHSRGFAGS
                   QGWQVEVPVADLNGMLSAKPISLKGSATLNQNVLLTVPDLQIKYGENYLKASGVLDDH
                   SDFALDINAPNLRGLWSDLKGRVKGRVAISGQITTPNLDLDLTSSNLHQGFQLAKAS
                   IKGHINNASLSSGKLNIKAEQLHYGGNIKLHLDLDLSGDEQNHKLILKSQGEPVAAN
                   LQINGHFDRTLEQWKGIISQVKFETPIGDVKSNQAIAVSYDNKQTQANIASHCWQNTD
                   VELCFPQAFNAGKQGNIPFQFKRVNLDLVNKLIEQNSLKGNLQOVGNVAWFTDKPFQF
                   TANVDGNHLAFSQKLDYRTFKLYIPKLTLNADIQNNNLVLKTDINVHNQGRIVGDIHL
                   NDLAKNRQLGGTLAIERLNLSIANQLLTSGESVNGEVVSKLSFGGNLEKPLLNGDFNI
                   RNIRTKLKSMPVNTDGDIALRFNDNRSTLQGKIKTVDSHLNLJIGRANWANIEHWTTE
                   LNAQANNFNVDIPSMAKLRFSPNITIKANPKELNLSGTVDIPWARIKIDSLPDTAEPV
                   SEDEVILNGPHKSKEELIKREFAAKTKSGMEIRSDLRINIGKDVSLDAYGLKTNLDGL
                   LSVKQDKGNLGLFGQINLTKGRYASFGQDLLIRKGLISFSGQATQPTLNEAIRNPET
                   MEDSKITAGVRVIGIADSPEVTIFSEPSKPQOQALSYLLTGRSLESSGEVGSTGSVGA
                   ALIGLGISKSGKLVGSIGEVFGIQDLNLGTSGVGDKSKVTVSGNITNRLQIKYGVLF
                   DGLAEVTLRYRLMPQLYFQSVSSTNQVFDLLYKFEF"
gene               complement(9994..11730)
```

```
alignment_scores:
          Quality:           598.00        Length:          201
            Ratio:             3.691          Gaps:            2
Percent similarity:          80.597   Percent identity:    59.204

alignment_block:
US-09-252-691-7056 x U32752

Align seg 1/1  to: U32752   from: 1  to: 11833
```

```
   15 LysThrSerPheArgLysHisArgValGluArgPheSerArgGlnAl  31
      |||:::::  ||||||| :::    :::   ||
 4342 AAAAGTGCGGTTAATTTCACCGACCACTTTAGCCAAATTAAAACCCAA 4391

   31 aThrArgArgThrProGluProGlnProThrArgValIleLeuPheAsnL  48
      |||       ||||:::||||||||||||
 4392 AACGTTGTCGGTTGGATGAA........ACCAAAGTGTGTTATTCAATA 4432

   48 ysProTyrAspValLeuProGlnPheThrAspGluAlaGlyArgSerThr  64
      ||||||||||| :::||||||||||||||||||||
 4433 AACCTTTGATGTACTCACGCAATTTACCGACGAACAAGGACGCGCCTACC 4482

   65 LeuLysAspPheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAs  81
      ||||||||||| :::  :::||||||||||||||||||
 4483 TTAAAAGATTTTATTCAATTCCAAATGTTATGCGGCCAGGTCGATTAGA 4532

   81 pArgAspSerGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnA  98
      |||||||||  ||| |||||||||||||||||
 4533 TAGAGACAGTGAGGATTACTGATTTTAACCAATAATGGCCAATTACAAC 4582

   98 laArgLeuThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGln 114
      |:::::::::  :::||| |||:::||||||
 4583 ATCGTTTGGACAGATCCCAAATTAAAACAGAAAAAAACCTATTGGGTGCAA 4632

  115 ValGluGlyGluProAspAspAlaSerLeuAlaLysLeuArgAsnGlyVa 131
      |||||||||  :::::||| ::::||||||||||||||
 4633 GTGGAAGGTATCCCTGAAGAAACCGATTTAGCACAACTGAAGAAAGGCGT 4682

  131 lThrLeuAsnAspGlyTyrProAlaGlyIleLeuGluArgValAsnG    148
      ||||||||| |||  |||||||||||
 4683 TGAACTCAAAGATGGCCTGACAAAATCCGCTAAAGTGCGGTTAATTTCAG 4732

  148 luProGluTrpProArgProAsnProProIleArgGluArgLysSer    164
      ||||||   ||||| ||||||||||||||||||||
 4733 AGCCAAAT...TTATGGGAAGCAATCCACCTATCCGAGAACGGAAAAT 4779

  165 IleProThrSerTrpLeuLysIleThrLeuTyrGluGlyTyrArgAsnArg1 181
      ||||||||||||:::||| :::  ||||||||||||
 4780 ATTCCCACAAGCTGGTTAGAAATTAAGATTAGTGAAGGGCCAACCGCCA 4829

  181 nValArgArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleA 198
      |||||||||||||:::||||||||||||||||||||||
 4830 AGTGCGGAGAATGACTGCGCATATTGGTTTCCCCACTTTACCGATTAGTGA 4879

  198 rgTyrAlaMetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrp 214
      :::||||| |||||||||||:::||| |||||||||
 4880 GAGTGAGTATGGGCTTATTATCTATTAATGGATTAGAAAAATGGCTCATTT 4929

  215 Arg 215
      |||
 4930 AGA 4932
```

seq_name: gb_ba2:SLE250721

```
seq_documentation_block:
LOCUS       SLE250721    5476 bp    DNA              BCT        07-JAN-2000
DEFINITION  Synechococcus leopoliensis crtQ gene (partial), dxr gene, ORF2,
            ORF3, ORF4 and ORF5 (partial), strain SAUG 1402-1.
ACCESSION   AJ250721
VERSION     AJ250721.1    GI:6689328
KEYWORDS    crtQ gene; deoxyxylulose 5-phosphate reductoisomerase; dxr gene;
            ORF2; ORF3; ORF4; ORF5; zeta carotene desaturase.
SOURCE      Synechococcus leopoliensis
  ORGANISM  Synechococcus leopoliensis
            Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
REFERENCE   1  (bases 1 to 5476)
  AUTHORS   Miller,B.H., Heuser,T. and Zimmer,W.L.
  TITLE     Characterization of the functional involvement of adeoxyxylulose
            5-phosphate reductoisomerase gene harbouringlocus of the
            Synechococcus leopoliensis genome in isoprenoidbiosynthesis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5476)
```

```
AUTHORS   Zimmer,W.
TITLE     Direct Submission
JOURNAL   Submitted (29-OCT-1999) Zimmer W., Atmosphaerische Umweltforschung,
          Fraunhofer Institut, Kreuzeckbahnstrasse 19,
          Garmisch-Partenkirchen, D-82467, GERMANY
FEATURES             Location/Qualifiers
     source          1..5476
                     /organism="Synechococcus leopoliensis"
                     /strain="SAUG 1402-1"
                     /db_xref="taxon:32047"
                     /country="Germany"
     gene            complement(1..856)
                     /gene="crtQ"
     CDS             complement(<1..856)
                     /gene="crtQ"
                     /codon_start=1
                     /transl_table=11
                     /product="zeta carotene desaturase"
                     /protein_id="CAB65434.1"
                     /db_xref="GI:6689329"
                     /translation="MRVAIVGAGLAGLAAAIDLVDAGHQVAIYDSRPFVGGKVGSWID
                     ADGNHIEMGLHVFFNYANLFALMRKVGAFENLLPKAHTHTFINKGGEVGELDFRFPI
                     GAPFNGLKAFFTTSQLTWLDKLQNALAGTSPLVRGILDYEGAMKIIRALDRISFADW
                     FRSHGGSEGSLKRMWNPIAYALGFIDTENISARCMLTVFQMFAAKTEASKLNLLAGSP
                     AEYLHKPILDYIQARGATLHLRRRVREIEYTETNGQTVVTGLQIADGDAVERVEADVY
                     LAACDVPGI"
     gene            988..2207
                     /gene="dxr"
     RBS             988..992
                     /gene="dxr"
     CDS             999..2207
                     /gene="dxr"
                     /function="involved in isoprenoid biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="deoxyxylulose 5-phosphate reductoisomerase"
                     /protein_id="CAB65435.1"
                     /db_xref="GI:6689330"
                     /translation="MKAVTLLGSTGSIGTQTLDILEQYPDRFRLVGLAAGRNVALLSE
                     QIRRHRPEIVAIQDAAQLSELQAIAIDLDNPLLITGEAGVTEVARYGDAEIVVTGIV
                     GCAGLLPTIAAIEAGKDIALANKETLIAAGPVVLPLLQKHGVITPADSEHSAIFQCI
                     QGLSTHADFRPAQVVAGLRRILLTASGGAFRDWPVERLSQVTVADALKHPNWSMGRKI
                     TVDSATLMNKGLEVIEAHYLFGLDYDIVIHPQSIIHSLIELEDTSVLAQLGWPDM
                     RLPLLYALSWPDRLSTOWSALDLVKAGSLEFREPDHAKYPCMDLAYAAGRKGGTMPAV
                     LNAANEQAVALFLEEQIHFSDIPRLIERACDRHQTEWQQQPSLDDILAYDAWARQFVQ
                     ASYQSLESVV"
     RBS             2206..2209
                     /note="ORF2"
     CDS             2217..2816
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB65436.1"
                     /db_xref="GI:6689331"
                     /translation="MYRYLLFHKPYDAVCQFSPSDRPDQQTLKDYIDVPEVYPVGRL
                     DRDSEGLLLILTNNGALQHRLCHPRFGHDRTYWVGVEREPTEAALQALRQGVQIQDYRT
                     RPAKYQRLDDDPQIPERDPPIRFRKTVPTAWLALTLQEGRNRQVRMTAAVGHPTLRLI
                     RAAIATHPDQPPLQLQGLAPGTWRDLTAIECRRLESLLR"
     RBS             2827..2830
                     /note="ORF3"
     CDS             2846..3757
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB65437.1"
                     /db_xref="GI:6689332"
                     /translation="MTAGMSESIPESEASFVAAADERTAMLSLSLAQMEVAATTDVGM
                     RRSQNEDFYVYHYQLEHLEPSGTSVQAKGLFVLCDGMGHARGEIASLEAARVFKQE
                     IEAQWTELPSREALRSAVLAANQAFELNQAAVSGSDRMGTTLVAILLQGTRVAII
                     HVGDSRAYRYSRLITGLDQLTRDHEVGQLAIAQGVDPDVAYARPEAHQLTQLGPRNND
                     WLDPEVRFLEVREDTVFLLCSDGLSDFDLVEEVTPTHVAPLLQTRTSLQYGIQQLVEI"
```

CDS
```
                ANEAGGHDNITVAVRLRLRPQLGPVI"
                complement(3791..4384)
                /note="ORF4"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAB65438.1"
                /db_xref="GI:6689333"
                /translation="MLGRQDQPLTADGLTQAEQLGKQWWQQTCCWRPQAIWTSPLQRAQ
                QTAIALSQGQTPVIPLLVQPALTEIDLGCLTGLTWPEARQRYPDLCQQLLRDRHWRPI
                PEAETLSDCRMRAEQVCTQLIHTEADLWVISHGGFLQHLLAALLGCDRSWGLDWPLLG
                AIDLSLAHEDWDTAGRDRWNTSLWRWQILHWPSRSAI"
```

RBS
```
                complement(4398..4402)
                /note="ORF4"
```

CDS
```
                complement(4410..>5476)
                /note="ORF5"
                /codon_start=3
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAB65439.1"
                /db_xref="GI:6689334"
                /translation="ACRSLLQEPFCLVLSDRLSFLVILVPYPSQPSCYFSQTPEVLAE
                AVRLLAQHTAHHFPALQPELARALHFRPLANVAETWAPFWAAVLQQWPTVAASDRPS
                TEEVSDRPDRSTVGADLELLQAIAHEVKTPLATIRTLRSLLRRVDLPAQARKRLEAI
                DLECSEQIDRFGLEFRAVELENRESQALLLARTSLTEVPAETVPWRKQAERRSLALD
                LQLPERMPAVVSDPTMLDQALTSLVERFTRNLPAGSRIEIVVSLAGAQLKLRLTATEL
                VSDPSEQPRTAPRSLGQLLTVQPETGSLSKLEVSKNLFQALGGKLTVRRRSPQASEV
                ITVFLPVEPSDDATTADSPR"
```

```
 ●ASE COUNT   1135 a    1592 c    1626 g    1123 t
 ORIGIN
```

alignment_scores:
```
      Quality:      475.50              Length:      188
      Ratio:          3.372             Gaps:          3
Percent Similarity:  75.000   Percent Identity:  53.723
```

alignment_block:
```
US-09-252-691-7056 x SLE250721
```

Align seg 1/1   to: SLE250721   from: 1   to: 5476

```
.42  ArgValIleLeuPheAsnLysProTyrAspValLeuProGlnPhe......  56
      ||| :|||||||||||::||||||:::::::::        ||||||
2226  CGCTATCTGCTGTTCAACAGCCCACGATGCTGTTGCCAGTTCAGTCC  2275

 57  .ThrAspGluAlaGlyArgSerThrLeuLysAspPheIleProValGlnG  73
      :::||:::     :::||||||||||||::||      ||||
2276  AAGGGGATCGCCCGGATCAGCAGACCCTGAAGGACTACATCGATGTGCCTG  2325

 73  lyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuVal  89
      |||||||::|||||||||||||||||||||||||||||||||:::
2326  AGGTCTATCCGTCGGTCGCCCTCGATCGCGATAGTGAGGGGTTGTTACTA  2375

 90  LeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysAr  106
      |||||||||||:||||| :||| ||||| ||:|:|
2376  CTGACCAACAATGGGCCACTGCAGCATCGTCTTTGCCATCCGCGCTTTGG  2425

106  gThrGlyLysIleTyrTyrValGlnValGlyGluGlyLeuProAspAspPheGluAs  123
      ::: ||||||||||||::|||| |||||:|||        :::||:
2426  TCACGGACGCCACCTATTGGGTGCAAGTGCAAGCGGGAACCGACCGAGGCAG  2475

123  erLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeu  139
      ::||::|||| :::::||| ::::|||| |
2476  CGCTCGCAGGCTCGCGTCAAGGAGTGCAAATTCAGGATTATCGAACCCGC  2525

140  ProAlaGlyIleGluArgValAsnGluProGluTrpProTrpProProArgAs  156
      |||||||||||||||:| ::: |||:|
2526  CCCGCCAAAGTGCAGCGATTGGACGACCCTCAG...ATTCCCGAGCGGGA  2572

156  nProProIleArgGluArgLysSerIleProThrSerTrpLeuLysIleT  173
      :|||||||||     ||||||||||||||||:||||| ::|
2573  CCCCCCGATTCCCTTCCGCAAAAACCGTCCCGACGACGGCATGGCTGGCCCTGA  2622
```

```
173  hrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisVal  189
      ||||   ||||||||||||||||||||||||||||||||||||||||||
2623  CTCTACAGGAAGGACGGAACGCCAAGTACGGCGGATGACCGCTGCAGTG  2672

190  GlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyr......  204
      ||||||||||||||||||||||||||||||||||||||||||||
2673  GGCCATCCAACGCTACGCTTGATACGACGCAGCAGCCATCCGTACCCATCCCGA  2722

205  ..........ThrLeuAspSerLeuAlaAsnGlyGluTrpProArgAspV  217
                    ||||||:::|||||||||||||||||||||
2723  TCAACCACCCTTCAGTTGCAGGGGCTCGCCCCGAACCTGGCGAGACT  2772

217  alThrProLysGlu  221
      ::|||   |||
2773  TAACGGCGATCGAG  2786
```

seq_name: gb_bal:D90899

```
COMMENT      Potential protein coding regions were assigned on the basis of
             similarity search of the ORFs and GeneMark analysis.
FEATURES             Location/Qualifiers
     source          1..133859
                     /organism="Synechocystis sp."
                     /strain="PCC6803"
                     /db_xref="taxon:1143"
     gene            1..772
                     /gene="ispB"
     CDS             <1..772
                     /gene="ispB"
                     /note="ORF_ID:slr0611"
                     /codon_start=2
                     /transl_table=11
                     /product="prenyltransferase"
                     /protein_id="BAA16579.1"
                     /db_xref="GI:1651651"
                     /translation="ARHRRLAEITEMIHTASLVHDDVVDEADLRRNVPTVNSLFDNRV
                     AVLAGDFLFAQSSWYLANLDNLEVVKLLSEVIRDFAEGEILQSINRFDTDTDLEYLE
                     KSYFKTASLIANSAKAAGVLSDAPRDVCDHLYEYGKHLGLAFQIVDDILDFTSPTEVL
                     GKPAGSDLISGNITAPALFAMEKYPLLGKLIEREFAQAGDLEQALELVEQGDGIRRSR
                     ELAANQAQLARQHLSVLEMSAPRESSLLELVDYVLGRLH"
     CDS             937..1494
                     /note="ORF_ID:slr0612"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16580.1"
                     /db_xref="GI:1651652"
                     /translation="MGRLDQDSEGLLLLTSNGKLQHRLAHREFAHQRTYFAQVEGSPT
                     DEDLEPLRRGITFADYPTRPAIAKIITEPDFPPRNPPIRYRASIPTSWLSITLTEGRN
                     RQVRRMTAAVGFPTLRLVRVQIVGTGRSPQQQKGKSAATWCLITLEGLSPGQWRPLTPW
                     EENFCQQLLTGNPNGPWQKKFGDRR"
     CDS             1577..2098
                     /note="ORF_ID:slr0613"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16581.1"
                     /db_xref="GI:1651653"
                     /translation="MSYLIAVVANRIAAEEAVTTLEQAGFAQKNLTIIGTGYKTADEF
                     GLVDPKKQAIKRAKLMAIWLVPFGFAAGYCFNLITGLSTLDWAGDPGNHIVGGLLGAI
                     GGTMGSFVVGGVGLSFGSGDSLPYRNLLQAGKYLVVVAGGELQKQRATNLLRPLNPE
                     YLQGYTAPDEAFV"
     CDS             complement(2172..2873)
                     /note="ORF_ID:sll0558"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16582.1"
                     /db_xref="GI:1651654"
                     /translation="MSDNLTELSQQLHDASEKKQLTAIAALAEMGEGGQGILLDYLAK
                     NVPLEKPVIAVGNVYQTLRNLEOETIITTQLQRNYPTGIFPLQSAQGIDYLPLQEALGS
                     QDFETADEITRDKLCELAGPGASQROWLYFFTEVEKFPALDLHTINALWLHSNGNFGF
                     SVQRRLWLASGKEFTKLWPKIGWKSGNVWTRWPKGFTWDLSAPQGHLPLLNQLRGVRV
                     AESLYRHPVWSQYGW"
     gene            complement(3192..4268)
                     /gene="PNIL34"
     CDS             complement(3192..4268)
                     /gene="PNIL34"
                     /note="ORF_ID:sll1214"
                     /codon_start=1
                     /transl_table=11
                     /product="PNIL34"
                     /protein_id="BAA16583.1"
                     /db_xref="GI:1651655"
                     /translation="MVNTLEKPGFDEIRPGVKTPAKETILTPRFYTTDFDEMAKMDIS
                     PNEDELRAILEEFRVDYNRHHFVRNESFNKSWDHIDGEKRQLFVEFLERSCTAEFSGF
                     LLYKELGRRLKNKNPLLAECFWIAEFRDEARHAGFLNKAMSDFNLSLDLGFLTKSRKYT
                     FFKPKFIFYATYLSEKIGYWRYITIYRHLEKNPDCIYPIFEFENWCQDENRHGDFF
                     DAIMRAQPHTLNDWKAKLWCRFFLLSVFATMYLNDTQRADFYACLGLEARSYDKEVIE
                     KTNETAGRVFPIILDVNNPEFYNRLETCVSNNEQLRAIDASGAPGVIKALRKLPIFAS
                     NGWQFIKLYLMKPIAVDQLAGAVR"
```

Right column:

```
     CDS             complement(4451..5389)
                     /note="ORF_ID:sll1213"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAA16584.1"
                     /db_xref="GI:1651656"
                     /translation="MSLSENQRILVTGGAGFLGKQVVAQLIAAGGDRPKITIPRSKDC
                     DLRQASACERAVENQDIVIHLAAHVGGIGLNREKPAELFYDNLMGVQLIHAAHQAGV
                     KKFVCVGTICAYPKFTPVPFKEEDLWNGYPEETNAPYGVAKKALLVQLESYRLQYGFN
                     GIYLLPVNLYGPEDNFDPRSSHVIPALIHKVYEAQKAGKQLPWGDGSPTREFLYST
                     DAARGIVMGTQAYDKADPVNLGTNFEISIKDLTELICELMEFEGDIIWETDQPNGQPR
                     RCLDTTKAKAEFGFEAQVSLREGLKNTIDWYRQQRS"
     gene            complement(5534..56511)
                     /gene="rfbD"
     CDS             complement(5534..6622)
                     /gene="rfbD"
                     /note="ORF_ID:sll1212"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-D-mannose dehydratase"
                     /protein_id="BAA16585.1"
                     /db_xref="GI:1651657"
                     /translation="MSKSKVVLLTGITGQDGSYLSELLLEKGYQVHGIRRTSFNTD
                     RIDHLVVDPHDLEAKLRLHYGDLTDGTTLRRILEDVKPTEIYNLGAQSHVRVSFDSPE
                     YTVDSVAMGTLRLLEAIRDYQHRTGIQVRFYQAGSSEMFGKVQEIPQKETTPFYPRSP
                     YACAKVYGHWQTVNYRESYDLFACNGILFNHESPRRGETFVTRKITRAIARIVAGTQK
                     KLYLGNIDSKRDWGYAKDYVRAMWAMLQOEQPDDYVVATGETHEVKEFLEIAFGYVNL
                     NWQNYVAFDERYLRPAEVDLLIGDPAKTKAQLGWEPSVITTELVHLMVEADLAVLGLT
                     SPNQSGRIKELMAQDMAFIRSQNGHAVD"
     gene            7229..8311
                     /gene="psbA2"
     CDS             7229..8311
                     /gene="psbA2"
                     /note="ORF_ID:slr1311"
                     /codon_start=1
                     /transl_table=11
                     /product="photosystem II D1 protein"
                     /protein_id="BAA16586.1"
                     /db_xref="GI:1651658"
                     /translation="MTTLQORESASLWEOFCQWVTSTNNRIYVGWFGTLMIPTLLTA
                     TTCFIIAFIAAPPVDIDGIREPVAGSLLYGNNIISGAVVPSSNAIGLHFYPIWEAASL
                     DEWLYNGGPYQLVVHFLGIFCYMGRQWELSYRLGMRPWICVAYSAVVSAATAVFLI
                     YPIGQGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSLFSAMHGSL
                     VTSSLVRETTEVESQNYGYKFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLG
                     AWPVIGIWFTAMGVSTMAFNLNGFNPNQSILDSQGRVIGTWADVLNRANIGFEVMHER
                     NAHNFPLDLASGEQAPVALTAPAVNG"
     gene            8492..10471
                     /gene="speA"
     CDS             8492..10471
                     /gene="speA"
                     /note="ORF_ID:slr1312"
                     /codon_start=1
                     /transl_table=11
                     /product="arginine decarboxylase"
                     /protein_id="BAA16587.1"
                     /db_xref="GI:1651659"
                     /translation="MGEPVPADKALGKKFKKKNASWSIEESEALLYRVEAWGAPYFAI
                     NAAGNTVSPNGDRGGSLDLELVEALRQRKLGLPLLIRFSDILADRLERLNSCFAKA
                     IARYNPNTYQAVYPVKCNQQRHLVEALVRFGQTSQCGLEAGSKPELMIALATLPPPL
                     DRQDKHTKPLIICNGYKDQDYLETALLAKRLGHRPIIIIEQLRELEWLHISQQLNIK
                     PMLGVRARLSCQSLKSSEISSGNGDRAKLGLTMPDIVTVIHRLEENNCLDCLKMLHFH
                     LGTQVSDIALIKEAMREASQLYVELVLGAKMRYLNVGGGLAVDYDGSKTNYPASKNY
                     NMQNYANDIVAAIQDACELGQVSPPIIVSESGRAINAHQSVLVFDVLGSNQTGFSEPH
                     PPDENAHPLLKNLWECYETITAEQYQEQYHDALQKTEASSLNFGYLSLTERGQAEQ
                     IHWACCRKIFEITRQLEYIPEDFQALDKIMTDIYYNLSVFQSAPESWSLDQLFPILP
                     IHHLNEKPSQRYVLADLTCDSDGKIDRFIDLWDVKSVLEWHPLENDGNPYYLGMFLVG
                     AYQEIMGNLHNLFGDINVHIATTPQGYQIESVVRGDTMTEVLGYVQYDSDDLLEGLR
                     RHTELALSNGQITLEESRRLLEDYEQSLRRYTYLS"
     gene            complement(10622..12631)
                     /gene="lig"
     CDS             complement(10622..12631)
                     /gene="lig"
```

6-phosphogluconate dehydrogenase;
7-beta-(4-carbaxybutanamido)cephalosporanic acid acylase; DnaK
protein; GTP-binding protein Era: IMP dehydrogenase; NADH
dehydrogenase subunit 4; Nifs protein; RNA polymerase sigma factor;
acetyl-CoA carboxylase beta subunit; acriflavin resistance protein;
alpha-mannosidase; aminomethyltransferase; beta transducin-like
protein: cell division cycle protein; chemotaxis protein CheA; 35.6
kD protein; iron utilization protein; low molecular weight
phosphotyrosine protein phosphatase; oligopeptide transport system
permease protein; photosystem I subunit VII; photosystem II PsbI
protein; phycobilisome LCM core-membrane linker polypeptide;
putative; riboflavin synthase alpha chain; sensory transduction
histidine kinase; succinate-semialdehyde dehydrogenase (NADP+);
transcriptional regulatory protein; transposase; unidentified open
reading frame; uridine 5-monophosphate synthase; uroporphyrin-III
synthase; xylose repressor; ycf37 gene product.

SOURCE   Synechocystis sp. (strain:PCC6803) DNA.
  ORGANISM   Synechocystis sp.
                Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE   1 (bases 1 to 125469)
  AUTHORS   Tabata.S.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-AUG-1995) to the DDBJ/EMBL/GenBank databases. Satoshi
                Tabata, Kazusa DNA Research Institute, Gene Structure 2: 1532-3
                Yanauchino, Kisarazu, Chiba 292, Japan (E-mail:tabata@kazusa.or.jp,
                Tel:0438-52-3933, Fax:0438-52-3934)
REFERENCE   2 (bases 1 to 125469)
  AUTHORS   Kaneko.T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N.,
                Sugiura.M. and Tabata.S.
  TITLE   Sequence analysis of the genome of the unicellular cyanobacterium
                Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
                region from map positions 64% to 92% of the genome
  JOURNAL   DNA Res. 2 (4), 153-166 (1995)
  MEDLINE   96127529
REFERENCE   3 (bases 1 to 125469)
  AUTHORS   Kaneko.T., Sato.S., Kotani,H., Tanaka.A., Asamizu,E., Nakamura,Y.,
                Miyajima,N., Hirosawa.M., Sugiura,M., Sasamoto.S., Kimura,T.,
                Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
                Okumura,S., Shimpo,S., Takeuchi.C., Wada,T., Watanabe,A.,
                Yamada.M., Yasuda.M. and Tabata.S.
  TITLE   Sequence analysis of the genome of the unicellular cyanobacterium
                Synechocystis sp. strain PCC6803. II. Sequence determination of the
                entire genome and assignment of potential protein-coding regions
  JOURNAL   DNA Res. 3 (3), 109-136 (1996)
  MEDLINE   97061201
COMMENT   Potential protein coding regions were assigned on the basis of
                similarity search of the ORFs and GeneMark analysis.
FEATURES   Location/Qualifiers
     source   1..125469
                     /organism="Synechocystis sp."
                     /strain="PCC6803"
                     /db_xref="taxon:1143"
     CDS   complement(<1..85)
                     /note="ORF_ID:sll0944"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA10018.1"
                     /db_xref="GI:1673331"
                     /translation="MQCLGKIFLSPGTKPESKNLIVIFFTAD"
     CDS   231..1826
                     /note="ORF_ID:slr0320"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA10019.1"
                     /db_xref="GI:1001397"
                     /translation="MIDRLLYVRLPCNPIFPIGVIYLADHVHKCCPPATAQRIFDLGTI
                     PPLDFNRALDECIDEFQPTILVFSWRDIOIYAPVGGRGGNPLQNAFEFYGKNPLVKL
                     RGALGGLKVTSAYYGELWNRLRLINRGLRRAKRYCSDPQIIVGGAGVSVFYEQLKTKL
                     PAGTIVSVGGETLLEKYLRQOTIEDERCIVGRSQPRPRLIHEQPSPMVKTACDYDY
                     IEQIWPAFDYYLQEDDFYLGVQTKRGCPHNCCYCVYTTVEGKQVRINPAEVVKEMRQ
                     LYDRGIRNFWFTDAQFIPARVFIDDVELLEAIAASGMEDIHWAAYIRADNLTPRLCE

alignment_scores:
  Quality: 474.50   Length: 210
  Ratio: 3.295   Gaps: 4
  Percent Similarity: 68.571   Percent Identity: 50.476

alignment_block:
US-09-252-691-7056 x D90899
Align seg 1/1 to: D90899 from: 1 to: 133859

  31 AlaThrArgArgThrProGluProGlnProThrArgValIleLeuPheAs 47
     ||| |||||:::||||:::||::||| ::|||:::|||
 808 GCGTTGAATAAAACTCCCCA...............ACCATTGTTTTTA 842

  47 nLysProTyrAspValLeuProGlnPheThrAspGluAlaGly......A 62
     ||||||||| |||||::| |||||||||:::|::|:::||| |
 843 CAAACCCTACGGAGTTCGTCGTCAATTTACCGATAATTCTGCCCATCCCC 892

  62 rgSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGly 78
     ::||| :::||| ::||| ::|||
 893 GGCCACGTTGAAGGATTATTTATTTGCCAGATTTATATCCCGTGGGG 942

  79 ArgLeuAsPArgAspSerGluGlyGluValLeuLeuThrAsnAspGlyVa 95
     |||||||||||||:::||||||| |||::||||||||::|:::|:::
 943 CGTTTGGATCAAGATAGCGAAGGACTATTGCTGCTCACCAACGGTAA 992

  95 lLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLysIleTyrT 112
     ||||| ||||||||:::   ::|||||::||||||::||| ||| |||
 993 ACTTCAGCATCGTTGGCCCACCGGAGTTGCCCACCAACGTACTTATT 1042

 112 yrValGlnValGlyGluProAspAspAlaSerLeuAlaLysLeuArg 128
     :::||| :::||||||| |||||| |||||||||
1043 TTGCCCAAGTAGAAGGCTCCCAACGGACGAAGAACCCCGCGG 1092

 129 AsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGluAr 145
     :::|||||||||| ||||| ::|||||||||:::|||:::
1093 CGGGGCATAACTTTCCGCGATTACCCTACCAGACCGGCATCGCCAAAT 1142

 145 gValAsnGluProGluTrpProLeuTrpProArgAsnProIleArgGluA 162
     :::||||||||||::|||||:::   :::::|||||||||||||||||||
1143 TATCACTGAACCAGATTTCCC...CCCAGAAATCCTCCCATTCGTTATC 1189

 162 rgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrGluGlyArg 178
     :::|||||| |||:::|||::|||||||||||||||
1190 GAGCCTCCATTCCCCACCAGTTGGTTAAGCATTACCCTAACGGAGGGGCGC 1239

 179 AsnArgGlnValArgArgMetThrAlaHisValGlyPheProThrLeuAr 195
     ||||||||||||||:::|||:::||:::||||||||||:::|||||||:::
1240 AATCGTCAGGTACGTCGAATGACACGCGGCAGTGGGCTTCCCTACCCTACG 1289

 195 gLeuIleArgTyr........................... 199
     |||:::|||||
1290 ATTGGTGCGGGTGCAAATACAGGTTACTGGTCGCTCTCCCAACAGGGCA 1339

 200 ....AlaMetGlySerTyrThrLeuAspSerLeuAlaAsn 211
              ||| |||:::::||:::|||:::|||:::||::|||
1340 AAGGTAAGTCAGCAGCAACTTGGTGCTTACCCTAGAAGGTTTGAGTCCG 1389

 212 GlyGluTrpArgAspValThrProLysGlu 221
     |||:::|||||| :::||:::
1390 GGGCAATGGGCGACCCCTGACCCCTTGGGAA 1419

seq_name: gb_bal:SYCSLRA

seq_documentation_block:
LOCUS   SYCSLRA   125469 bp   DNA   BCT   13-FEB-1999
DEFINITION Synechocystis sp. PCC6803 complete genome, 18/27, 2267260-2392728.
ACCESSION D63999 AB001339
VERSION   D63999.1 GI:1001396
KEYWORDS   era; ORF1; 1,4-alpha-glucan branching enzyme; 2-isopropylmalate
           synthase; 3-ketoacyl-acyl carrier protein reductase;

/note="ORF_ID:sll1209"

```
                LMVQTGMNYFEIGITSGSQELVRKMRMGYNLRTVLQNCRDLKGAGFNDLVSVNYSFNV
                IDETDTIRQTIAYHRELEAIFGADKVEPAIFFIGLQPHTHLETYALDKEILKPGYDP
                MSMMPWTAKKLLWNPEPLGSFFGEVCLQAWQQNPNDFGREVMNILEQRIGKADLETAL
                HSPLPDKKKFPPTMAEGKKLSPI"
gene            complement(1867..2904)
                /gene="rpoD"
CDS             complement(1867..2904)
                /gene="rpoD"
                /note="ORF_ID:sll0306"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase sigma factor"
                /protein_id="BAA10020.1"
                /db_xref="GI:1001398"
                /translation="MVTVTVILLFIKESFRMPTANLSSPTSPPTFTADMVRSYLHEI
                GRVPLLTHEQEIILGKQVQQMMALLEHKKALADRLGREPSDPWAEEAADLSVTKLHRY
                LGQGRAKRMIEANLRLVVAIAKKYQKRRNMEFLDLIQEGSLGLERGVEKFDPTKGYK
                FSTYAYWWIRQAITRAIAQQGRTIRLPIHITEKLNKIKTQRELSQQLGRSATPAEVA
                KALEIDPSOIREYLSLSRQPISLDVRVGDNQDTELSELLEDEGVSPDAYITQESMRQD
                LQNLLAELTPQQQAVLTMRFGLNDGQELSLAKIGQHLNISRERVRQLENQALAQLKRR
                RANMAEYIIAS"
gene            3365..4312
                /gene="era"
CDS             3365..4312
                /gene="era"
                /note="ORF_ID:slr0321"
                /codon_start=1
                /transl_table=11
                /product="GTP-binding protein Era"
                /protein_id="BAA10021.1"
                /db_xref="GI:1001399"
                /translation="MDIPNTTATIATIPQAPAGFRSGFVAIVGRPNVGKSTLMNQLVG
                QKIAITSPVAQTTRNRLQGIITTPSSQILLDTPGIHKPHHELGRVLVKNAIQAIHSV
                DLVVFLVDSSATLGRGDRFVVDLLQKTDGPVVVGLNKQDQQPPDQREELNASVETLTE
                NHGWPCFKFSALTGEGLSNFQSALEARLDDGPYYYPPDLVTDQPERFIMAELIREQIL
                LLTRQEVPRSVAIALEKVEETPERTNVFAAITVERGSQKGIIIGQKGSMLQAIGTAAR
                QQIOKLISGDVYLKLFVKYEPKWRQSRQQLEFGYRVEE"
gene            4322..7609
                /gene="cheA"
CDS             4322..7609
                /gene="cheA"
                /note="ORF_ID:slr0322"
                /codon_start=1
                /transl_table=11
                /product="chemotaxis protein CheA"
                /protein_id="BAA10022.1"
                /db_xref="GI:1001400"
                /translation="MTSDPNPAQAFENWFEEEGDDAFTSLIAGEETNSGIEELDFGD
                LFEGEPTVIQGEGFGSPSSSGEEDPPVRMTLDFNSPEAELDALFTDEDAPTTAIDPFES
                SSWLEIDQMMEADPLASIPETNAPAYQTAGGLDDNDLAELQDLLGVGDVEPAITNPGL
                HSFAGADRVTPLTNEVLLFAAIDDLDGLINAPVLDEEFSFTELETLIDTPAPKSFATI
                SHPFRLPVAMEEDEDEFKDLEALLEQADSSLGGGIAGMGGARLPARPRQVKSRSFEQT
                MRVPIKQLDNLSNLIGELVVKRNRLEQDQERLRQFLDNLLNQAQNLSDVGGRMQDLYE
                KNGLELLSELQADEELSSIPVALLTSRGADRHRKVAAKLGATGYFTKPYTEKDILDAA
                NRMLNGEVLLPGSTRKTNSLSRFKKARGNFSEEAPTIVRSNPGTLRSTVLIIDDSVM
                VREMLSMTFKKADYHTETARDGQDAWEKLKSGLPCDLILCDIEMPRMNGLELLERIQE
                DPNLSRIPVAMITSRGAQKMQRMAAEKGAKGYFVKPYIEETLLEASQRLIRGEVLLDL
                AALSADE"
gene            7609..10737
                /gene="ams1"
CDS             7609..10737
                /gene="ams1"
                /note="ORF_ID:slr0323"
                /codon_start=1
```

```
                /transl_table=11
                /product="alpha-mannosidase"
                /protein_id="BAA10023.1"
                /db_xref="GI:1001401"
                /translation="MTESLSPLLPEISIALVLEQLRRLCRQNCQSHWYCLPGAPWEIS
                DWQTYPLGKVNEKGYLIWEKGEKTQWFAQTFTVPKALADYPLDGLTLRLALTWWAKDA
                QIFVNGELVQAGDLFDSKARVLLTDNSQIGRTFVALKLTSSGHDIGGLMQSQLIFER
                KYPDLDPGFIADEIEVLSLYLAQFESERLAQLSKTLEAITWNKISDQEQFDQSLLQLR
                ORLLLTEGLKQQQINLLGHAHLDMAWLWQLEETWEVGERTFQSVINLQTEFADLVFG
                HTSPVLYQWIEEHLPALFEQIQVAIKKGNWELLGGMWVEPEVNIISGESLARQFLYGQ
                RYFEAKFGQISKVAWLPDSFGFCGQLPQIFHQSGIDYFVTGKLHWNDTNSFPHGAFHW
                RSPDGTEIFTVMSPPNLAGVMDNQOPLPMANYANRWQEQTGLKECLWLPGVGDHGGGPS
                RDMWEMKERWQNSEFFPINTAKAEHFLAGIKQKLQSDQSFPIWERELVIELHRGCFT
                VHADQKLYNRQCEHLLYEAELWSSFANWLGGYSYPKEKLELAWKKVLLNQFHDILPGT
                SIPEVFVTANQAWQEVYQTTNKIINQALKHLSDQINYPSYHHDSTERETPLLVFNPLH
                WSQSQLVEIPIQEGKHYQVYDGQKTLVNSQITAEGSLIFLVDSPLTCKYYWLKTKE
                PLSNSQQLESFSQDPILNNGLLKVEISLDNGEIASIYDLANQRQVLAGHNQLQFFRD
                QGQYWDAWNIDPNYEQYPLSSAKLINYQWLEQGPLHWRLRVKKFQQSVFTQDYCLTK
                DSPLLEIKEVDVRVEGVDNQDTELSELLEDEGVSPDAYITQESMRQD
                VPHQHWFSLIDQTENYGVSLLNNCKYGCDVKDNLMRLTLLRSSVWPDPSADRGKHKFS
                YYFYPHQGNWQSAKTVAVGYGIHRPLRTYFPENFNPLSSQKLLSPGEPWLNLGSDNLC
                LMALKLREDYPDDQWVLRCYEMSGESAQLEVITAFDHRVKTRLNLLEEAQPEENNHVN
                PWQIVSVCLAYD"
gene            10730..11614
                /gene="appC"
CDS             10730..11614
                /gene="appC"
                /note="ORF_ID:slr0324"
                /codon_start=1
                /transl_table=11
                /product="oligopeptide transport system permease protein"
                /protein_id="BAA10024.1"
                /db_xref="GI:1001402"
                /translation="MTKTFPLSGGSVAINLQRKIFQKPLASLSLITLIIIVLSVVFGP
                IFYRVPIDQIDFSQTAVPPSWQHPLGTNDLGQDQLARLLIGGRISLAVGLSAMAVSLT
                LGTLIGAIAGYFGGWDVVLMRITDLFLALPQLPLVLLVWYLFREPVTRATSPEKGIF
                ILVVLLIGLLNWMSVARLVRGNILRLREMEFVRAAIAMGAGSNHIIWRHLLPNVLNLI
                IVAASLAVGNAIITESTLSFLGLGFPPDVPTWGQMLYTAKDYLETAPLMAVFPGLAIF
                ITVLSVNYLGDGIGKNS"
gene            11662..12147
                /gene="appC"
CDS             11662..12147
                /gene="appC"
                /note="ORF_ID:slr0325"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAA10025.1"
                /db_xref="GI:1001403"
                /translation="MIRPVPDASNQFYLDHINLLRQSYERLTGKTFGPSGLTGVALAQ
                AVYEAPYVVVSHGIEPDPIFNYANLTAQKLFEFPWVEFCQLPSRQSAELPNQTERQQL
```

```
alignment_scores:
     Quality:      298.00          Length:    187
       Ratio:       2.614            Gaps:      3
Percent Similarity:  60.963    Percent Identity:  38.503

alignment_block:
US-09-252-691-7056 x SYCSLRA

Align seg 1/1  to: SYCSLRA   from: 1  to: 125469

 44  IleLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAl  60
     ::|||  ||||||||  |||||||||||:||||||::||||  |||
78736  TTATTGGTTAATAAACCGAGGGATGTGCTGCCACCTGTCGACGATCCCG  78785

 60  aGlyArgSerThrLeuLysAspSPheIleProVal...............G  72
     ||||||  |||::||:||||  ::|:::||||                  :
78786  GGGCCCGAAAACAGTGCTGGATTTACTGCCCCAGGACTTACAGCGGGGCA  78835

 72  lnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeu  88
     ::|||::|||||  :::||||||||||||||:::|||  ||:  |||
78836  AAGGACTGCCATCCCGTTGGCCGGTTAGACCGTAATTCCACTGGCGCTCTC  78885

 89  ValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLyY  105
     :::||||||||||||  |||  |||||  |||  ||:  :::||:
```

```
78886  CTGCTCACCATGATGGCGAATTGACTCTGCGGCTGACCCATCCCGTTA 78935
105    sArgThrGlyLysIleTyrValGlnValGlyGluValGlyProAspAspA 122
       :::::      |||     |||     :::||||||:::||||||||||
78936  CCATTTGCCAAAAACCTATGACGTTTGGTTAGAAGTAATCCCAGTGATG 78985
122    laSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
       :::|||     ||||:||||:     ||:|||     |||
78986  AAGATCTGGAAAAGTGGCATCGGGCATGATGCTAGATGGCAAAAAAACC 79035
139    LeuProAlaGlyIleGluLeuArgValAlaAsnGluProGluTrpProroAr 155
       ||||||||     :::||||     :::::|||
79036  CTACCAGCTACTCTGGAAGTAATTCCGAAAATAA............... 79071
155    gAsnProProIleArgGluArgLysSerIleProThrSerTrpLeuLysI 172
                                     |||     |||
79072  ................GACCAGATTCATTGTTGG 79090
172    leThrLeuTyrGluGlyValArgAsnArgGlnValAlaArgMetThrAlaHis 188
       ::|||||     ||||||||||||||||:||:::::|||:::::|||
79091  TGACCTTGACGGAGGACGTAATCGCCAAATTCGTCGTTAGCCGAAGAA 79140
189    ValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrTh 205
       ::::|||:     :::||||:|||     |||     :::::::|||
79141  TGGGATTGACGGTGCTGAAGCTCCATCGCCGCCGCCACCATTGGCCCCTACA 79190
205    rLeuAspSer..........LeuAlaAsnGlyGluTrpArgAspValT 218
       ::::|||:::     ::::AGTTCTGGGTAGTGGGGCAATTTCGTTTTCTTT 79240
218    hrProLysGlu 221
       :::|||
79241  CCCCGGCGAA 79251
```

```
seq_name: gb_ba2:AE001942

seq_documentation_block:
LOCUS       AE001942    9999 bp    DNA    BCT    22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 79 of 229 of the complete
            chromosome 1.
ACCESSION   AE001942 AE000513
VERSION     AE001942.1  GI:6458611
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcus.
REFERENCE   1  (bases 1 to 9999)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 9999)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr. Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..9999
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            complement(71..655)
                     /gene="DR0891"
     CDS             complement(71..655)
                     /gene="DR0891"
                     /note="similar to GP:1498492 percent identity: 61.93;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-binding response regulator"
                     /protein_id="AAF10469.1"
                     /db_xref="GI:6458612"
                     /translation="MRLVIADDHPLFRMGLKYALIQQGFDVAEAADGPSALDACRTY
                     QPDAALLDVKMPGMTGIEVCDRLRLTAPRVVSLITTFAEPAIVQAARAGAKGYVSK
                     ETDPESLARQLREIVAHPEVDRLPQVDVPRLTPRESEVLPLLARGFSNKEIARDLGVS
                     PDTVKDHLARLYSKLNAGDRTEAVSRARSIGLLS"
     gene            complement(652..1761)
                     /gene="DR0892"
     CDS             complement(652..1761)
                     /gene="DR0892"
                     /note="similar to PID:245589 percent identity: 47.14;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sensor histidine kinase"
                     /protein_id="AAF10470.1"
                     /db_xref="GI:6458613"
                     /translation="MTRMTPPAAPRAAAAPPERRPARLAAVRAGRSRPPLQTLQTQFT
                     LVVSLLAFLPNLVLALASGAGSLSWSAAAWMLLVGLLSAVIGYVLSGALLRPLHRLEL
                     EVSEGDFAQFHADDPREIRSLRRAFGSLMTRLGETEQTRRNAFMATLVHDLKTPLIATG
                     HLTRALTELPLPDDEKRAVGQEIQTETERLALVQQMADAHRFERDDVQLHLAPGDLR
                     EVAEQAARRLRPQAEARGLSLHVSGSGQAEIDAAALERAVGNLTENALRYARHEVRLT
                     AAPGLLRVQDDGPGLSASLDDLAQPFNAQPAVIAGQQYTAGTAGLGLFIVRRIAEAHG
                     GTLRYGREPPRGRLPAAPGAFPLTFTLVLPEVLP"
     gene            complement(1922..2617)
                     /gene="DR0893"
     CDS             complement(1922..2617)
                     /gene="DR0893"
                     /note="similar to GB:U00096 PID:1787004 percent identity:
                     66.98; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF10471.1"
                     /db_xref="GI:6458614"
                     /translation="MVKSMQQIAMTQQKTLDQVRTFMARTYSWMAAGLALTAGVAYLT
                     AQNEGLAMQVASLRLPLMLAQLALVFVLSMFAQRLSAAVAGALFVGYAALTGLTFSAL
                     LFAYSPAAVITAFAVSAGTFGLMSVAGFVIKKDLSAMGRFFLFAVLGLVVAMLVNLFV
                     GSSALSLGSMIGVFLFAGLTAYDFQMLRNLALSGISGEQAERASINGALALYLDFIN
                     IFLFLLNIGNSRD"
     gene            complement(2702..4207)
                     /gene="DR0894"
     CDS             complement(2702..4207)
                     /gene="DR0894"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF10477.1"
                     /db_xref="GI:6458620"
                     /translation="MIKGKELLGRPVVAVSNGAQVDNVHDVVFDHQGNRVLALLVEEG
                     GWLRAAKAVPFGRVQSIGEKAVMIATPEDVVNTGRDPVLKEALAGKTNLIGMTLLTTD
                     GQTLGKITDVYFDERSGQVEGYEASGGLFADMTNGRTFIPAPQSIQIGRKDSAIVPAGV
                     AAAMREQAPGGIRGAFHTRAGKSLGGAVQGADSVRGAYEDTSRASVSRENYEQVSANVAE
                     AAREGQQQARAAASDLRSRVQGGAEDLREASAERQKDFYVGKVAGQDVRTDNGVLIVA
                     KGTTIDEGQARFAEQNGALPALLAAAGGSVREGVRDLLADPSSDAGDVVGRRVQEDV
                     RSSVGSLIAVQGQIVTPEVAERAREGLEERLVATQAPGLQTPPVDRAELREQASD
                     LRDNLADGLSNVSQNVSQGANELLGRAKSWLGERREDLEDSLDRQEEAYQARIRDAL
                     GRPVTRAVLTRTDRVILQPGELVTHAALDEARAAGVLGLLLDSVDGGAPLS"
     gene            4311..4697
                     /gene="DR0895"
     CDS             4311..4697
```

ARTURO REVERTE

```
             211 snGlyTrpArgAspValThrProLysGlu 221
                 ||||||:||||:||||:  :  |||:||:::::
            5333 CCGGCGAATACCGCGAACTCGGCCCGCGTGAC 5364

seq_name: gb_ba2:AE001708

seq_documentation_block:
LOCUS       AE001708    19531 bp    DNA          BCT       02-JUN-1999
DEFINITION  Thermotoga maritima section 20 of 136 of the complete genome.
ACCESSION   AE001708 AE000512
VERSION     AE001708.1  GI:4980740
KEYWORDS
SOURCE      Thermotoga maritima.
  ORGANISM  Thermotoga maritima
            Bacteria; Thermotogales; Thermotoga.
REFERENCE   1  (bases 1 to 19531)
  AUTHORS   Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
            Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
            McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
            Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
            Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
            White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Evidence for lateral gene transfer between Archaea and bacteria
            from genome sequence of Thermotoga maritima
  JOURNAL   Nature 399 (6734), 323-329 (1999)
  MEDLINE   99287316
REFERENCE   2  (bases 1 to 19531)
  AUTHORS   Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
            Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
            McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
            Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
            Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
            White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..19531
                     /organism="Thermotoga maritima"
                     /db_xref="taxon:2336"
     gene            104..226
                     /gene="TM0242"
     CDS             104..226
                     /gene="TM0242"
                     /note="similar to percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35334.1"
                     /db_xref="GI:4980742"
                     /translation="MRLEAKQSIPEFLNGFEISVEGSRVRLRKQEREYEFEVEK"
     gene            223..1074
                     /gene="TM0243"
     CDS             223..1074
                     /gene="TM0243"
                     /note="similar to GB:Pyro_h percent identity: 58.74;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAD35334.1"
                     /db_xref="GI:4980742"
                     /translation="MMLRASYGTIRKVNGNSSLEMDTAYLMLGERCVYNCLYCAQSRS
                     STSPSHFLSRVTWKEVSAELGGLKLNDRFKRVCIQVVSIPSYGKDLRLJIPRFRIPVSV
                     SVRAVSIEEVVEYPELGADRVGIAVDVPNKTLFEKIRGGKYERHLHILEKVSEMFPGR
                     ITHVIVGLGESDKIVDFTWARERNIVVSLFAFTPIKGTAFENRERPSLERYRKIQ
                     LVTYLLEKNLIKPENIIFDSNGKIIDVEWNGEFPEEALRTRGCPHCTRPYYNESPRGP
                     IYNVHWR"
     gene            1031..2386
                     /gene="TM0244"
     CDS             1031..2386
                     /gene="TM0244"
```

```
                     /note="similar to GB:X72888 GB:X79064 PID:435525
                     PID:483572 PID:1905810 percent identity: 60.89; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="electron transport complex protein, putative"
                     /protein_id="AAD35335.1"
                     /db_xref="GI:4980743"
                     /translation="MKAQEVPSTTFTGGETVLFTKGGVHPPELEKWSKDKPIERAPLP
                     QKVFVELSNHAGNPAKPVVSPGDEVKTGVIGEPEGFISAYLHSPVTGRVLEIKKILH
                     PILGKPIEAIVIERTSDDEWVHIETGDFERMSKEEILEIIKKAGIVGLGGAMFPTHVK
                     LSPPPEKVDTLIVNGAECEPVLTIDHRLMLERAEDILQGILIMMKVLGQKAVVGVE
                     SNRMDAYHNLKKVFKGYPVDVALLRTKYPQGAEKQLIYAITGRMVPRGGLPMDVGVV
                     QNVGTCVAVKEAVVDGKPLVERGMTVSGDAVKNQKNLIVRIGTPVKDVIDYCGGIDEN
                     TERVILGGPMMGISITNLDIPVMKGTSGITAFLPKKSRPQKPCIRCSECVQVCPMNLQ
                     PYLYLLSRKRKYDEAVENGLMDCIECGSCTYTCPSKIEHVRYIKLAKTVYRATRRGR
                     R"
     gene            2383..3339
                     /gene="TM0245"
     CDS             2383..3339
                     /gene="TM0245"
                     /note="similar to PID:663269 PID:663271 percent identity:
                     58.80; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="Na-translocating NADH-quinone reductase, Nqr2
                     subunit"
                     /protein_id="AAD35336.1"
                     /db_xref="GI:4980744"
                     /translation="MKLISAYAPHLREEDDVRKIMLDVLIALSPAVIGAAYFFGWYAL
                     FLCIAGAVIGELFDIFVMRYLRGVKDFVPDGSGAVTGLLLAMNVSTRLPFWAFLLGLV
                     FALGIGKHVFGGLGQNIFNPALVGRAFLLISFPTYMTTWVPGAGFWKSPADVVTAAT
                     PLALFKEHGVFTPYWDLFIGKVGGSLGETSALLLIGFIYLLLRKRVKIFIPVSYIGT
                     VLVFSSIAYLMNPRYGDPLFHLLSGGLMLGGALFMATDMVTSPITAKGQVIFGIGCGVL
                     TMAIRLFGAYPEGVSFSILFMNALVPLIDRYTRPRIFGEVKK"
     gene            3336..4010
                     /gene="TM0246"
     CDS             3336..4010
                     /gene="TM0246"
                     /note="similar to percent identity: 0.00; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAD35337.1"
                     /db_xref="GI:4980745"
                     /translation="MKDILKTGLIIMVFTAISGLFLGLVVGVKGKIQEADNAAKLSA
                     IKFVLKDPLTGDYLVDEKEIEEIVKKTGIETVVLKEYKEGVLGPLYEFVTKDGRNAY
                     VLSGYAPGFGGNVTVVACFINVIDYSQEFPMLNSVRVIDYSQEPGLGAKIGEESIQRRFFP
                     VPPEGLKNGLRVDKDAGLPKGSPEELKKQGIVKVSDVMTGATITPRAVVTALNLMRY
                     LEEYSK"
     gene            4007..4609
                     /gene="TM0247"
     CDS             4007..4609
                     /gene="TM0247"
                     /note="similar to PID:663269 PID:663273 PID:893414 percent
                     identity: 64.13; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="Na-translocating NADH-quinone reductase, Nqr4
                     subunit"
                     /protein_id="AAD35338.1"
                     /db_xref="GI:4980746"
                     /translation="MSRLRELTKGIIKENPTYVQVLGMCPTLAVTTSAINGLGMGLAT
                     TAVLTMSNVVISLIRKIVPDKIRIPIFIVVIASFVTMIDLLMHGFAYDLWKTLGLFIP
                     LIVVNCIIMGRAESFASKHGVLDSMLDGLGVGLGFTGSLVLLGSVRELEFGNGTIFGYK
                     VWELKIELPPGAYITLGLLSALFTYIGIRRKRGEAK"
     gene            4606..5181
                     /gene="TM0248"
     CDS             4606..5181
                     /gene="TM0248"
                     /note="similar to PID:893416 PID:893415 percent identity:
```

```
                                              |||||||||  ||||||                    ::  :::               |||
                                      17242 ATTCTCTTCACAAGCCTTCAGGATACGTGACCAGCACGAAAGATCACA 17291
68.11  identified by sequence similarity; putative"
/codon_start=1
/transl_table=11                            60 aGlyArgSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaA  77
/product="Na-translocating NADH-quinone reductase, Nqr5           ::::::   ::::  :::  ||:::: ::::: ::|||:::|||:|:::::
subunit"                              17292 TTCGGAAACCATCATGGAATTTCTCCCTCCATTGAAAGGAATTTCCCCG 17341
/protein_id="AAD35339.1"
/db_xref="GI:4980747"                       77 laGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsnAsp  93
/translation="MKVFLLFFSAIFVNNFVLARFLGICPFLGVSKRLETATGMGIAV       ::||||||||:|:::||:|||||||||||:::||||||||:|||||||||
TFVMTVSAISWFLDKLLISTGLEFLRTIVFIIVIASFVQFVELFLKKTSPDLYEALG 17342 TTGGAAGACTCGACAAAGATGCCGAGGTCTCCCGATAATAACGACGAT 17391
IFLFLITTNCAILGMVLLNSLMKLNFVEAVFHALGSGLGFALALVIFAGIREKMDLYD
LPEPFKGTAIALITAGLLSLAFMGFQGMVKL"            94 GlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLysIl 110
gene       5186.  .5488                      ::::|||||||:::|||     :::||
/gene="TM0249"                        17392 GGTGATTCGCTCATCGTGTGATTTCACCGAAATGGTCTGTAGAAAAGA 17441
CDS        5186.  .5488
/gene="TM0249"                             110 eTyrTyrValGlnValGlyGluProAspAspAlaSerLeuAlaLysL 127
/note="similar to GB:D13625 PID:216928 PID:435526            ||| ||:::|||:::||||||||||:: :||:::::::||:|:::
PID:1905809 percent identity: 64.86; identified by    17442 GTACATTGTGAAGGTAGAAGGAGAGTCACAGAAGACAAGATAGAAAAGT 17491
sequence similarity; putative"
/codon_start=1                             127 euArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIle 143
/transl_table=11                             ||:::||||||||||||||||:::||||||      :::   |||
/product="rnfB-related protein"       17492 TGAAAAATGGTGTGACTTTAAGGGATGGT.....TTTTTCGGCGAAAGCA 17535
/protein_id="AAD35340.1"
/db_xref="GI:4980748"                      144 GluArgValAsnGluProGluTrpLeuTrpProArgAsnProProIleAr 160
/translation="MKELEVIYSTLLLAVLGFGFGAFLAYSAQRFKVEDDPRVKMITE       ::::|||:|
VLPGINCGACGFAGCEAYAKAIVKGQAETNRCLFGRPQGVEEKIKKILEEYKNVSS"  17536 AAGCGCGTG..................................... 17544
gene       5475.  .6488
/gene="TM0250"                             160 gGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrGluG 177
CDS        5475.  .6488                        |||||||||||    ||:::|||||||  :::  ||||
/gene="TM0250"                        17545 .......GAGAGCTGTCGAACGACACGTTGAAGATAGTATAACGGAGG 17587
/note="similar to GB:L42023 SP:P43862 PID:1006171
PID:1221098 PID:603932 percent identity: 61.21; identified   177 lyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPheProThr 193
by sequence similarity; putative"            ||::: :::||:::||||||||||||||      ||||||:::  |||
/codon_start=1                        17588 GTAAGTATCATCAGATCAAACGAATGACAGCCGCCTGTGGGTCTAAAAACA 17637
/transl_table=11
/product="DNA processing chain A"          194 LeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeu...AspSerLe 209
/protein_id="AAD35341.1"                      ||||||||||:::||||||||||||||| ||:::::  |||||:::::
/db_xref="GI:4980749"                 17638 GTGCACTTGAAGAGAACAAGAATCGGTGGACTCGTTTACCCGATGACAT 17687
/translation="MSPLEMALLVHRGEFHLRELEPELPLEKFLKNADPKKTRKFLEK
CGKEELEROKELIRKHNVKLVSFWEDDYPQHLREIRYPPAVLFVRGDAELLKEKCVGV   209 uAlaAsnGlyGluTrpArgAspValThrProLysGlu 221
VGTRRPTSYGVNVRKRFVKLLSEYFVIVSGMAFGIDSVAHKEALSSGGKTVAVLGTGV       ||||:::||  ::::   ||||::||  :::|||
DVVYPRSNERLFHEIVKNGCVVSEYPMGTRARKHHFPARNRIIAGLSDAIIVTEADYR  17688 GAAACCCGGAGTACAGATTCTTGAGTGAAGAGGAG 17724
SGALITVKFALESGRDVFAVPGDIDRKTSEGTNYLIKSGAYPLTDEEDLETHFGIRRI
ASPSLDDDKKKIYDLLRSSPKTVDELVEELGWSVSEVLRVISEMELMGMIWFDGGAYR
LLG"                                  seq_name: gb_ba2:AE000741
gene       6493.  .6744
/gene="TM0251"              Length:      179   seg_documentation_block:
CDS        6493.  .6744                       LOCUS       AE000741      15077 bp       DNA                    BCT       25-MAR-1998
/gene="TM0251"                                DEFINITION  Aquifex aeolicus section 73 of 109 of the complete genome.
/note="similar to SP:P31803 percent identity: 78.33;   ACCESSION   AE000741 AE000657
identified by sequence similarity; putative"  VERSION     AE000741.1  GI:2983841
/codon_start=1                                KEYWORDS    .
/transl_table=11                              SOURCE      Aquifex aeolicus.
/product="carbon storage regulator"          ORGANISM    Aquifex aeolicus
/protein_id="AAD35354.1"                                  Bacteria; Aquificales; Aquificaceae; Aquifex.
/db_xref="GI:4980762"                         REFERENCE   1  (bases 1 to 15077)
/translation="MLVLIRRVGEKIVIGEDIVITVLKIEGNSVKIGIEAPKHVKILR     AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
EELYEELKSENIKASEVSKDDLKGVLRNDKGYKGPSASS"                            Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
gene       6707.  .6997                                             Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V.
/gene="TM0252"                                  TITLE     The complete genome of the hyperthermophilic bacterium Aquifex
CDS        6707.  .6997                                             aeolicus
/gene="TM0252"                                  JOURNAL   Nature 392 (6674), 353-358 (1998)
                                                MEDLINE   98196666
alignment_scores:                             REFERENCE   2  (bases 1 to 15077)
  Quality:      264.50          Length:      179     AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
  Ratio:        2.280           Gaps:          3                 Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
Percent Similarity:  64.804    Percent identity: 35.754           Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V.
                                                TITLE     Direct Submission
alignment_block:                                JOURNAL   Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,
US-09-252-691-7056 x AE001708                              CA 92121
Align seg 1/1  to: AE001708  from: 1  to: 19531  COMMENT     Putative indicates no similarity to known proteins
                                                          Hypothetical indicates similarity to a protein of unknown function.
       44 IleLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAl 60  FEATURES             Location/Qualifiers
                                                          source           1..15077
```

```
gene          /organism="Aquifex aeolicus"
              /strain="VF5"
              /db_xref="taxon:63363"
gene          complement(39..596)
              /gene="aq_1446"
CDS           complement(39..596)
              /gene="aq_1446"
              /codon_start=1
              /transl_table=11
              /product="putative protein"
              /protein_id="AAC07402.1"
              /db_xref="GI:2983848"
              /translation="MMKFLLILIFLASFSFSLTPEEKQLLKDIAEIKTTLKAFIREA
              DERFEQVDKRFEDVNRRFEDFNRRLNDLREDMNKRFELVDKRFIELREDMNKRFELVD
              QRFEQLYTFLWIITGIFTTLTASVIAFAWDRRTIIRKTKEETFEEMEKELKPEKFRK
              IMNALREKAKTDKELEAILKKYGLL"
gene          complement(608..2728)
              /gene="aq_1447"
CDS           complement(608..2728)
              /gene="aq_1447"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="AAC07406.1"
              /db_xref="GI:2983852"
              /translation="MGKRALLNLYRIEYRPKDTTFTVFKPTHEIQKEKLNKVRWRVFL
              QTGLPTFRREDEFWCAGKVEKDTLYLTLSNGEIVELKRVGEEEFRGFQNERECQELFR
              DFLTKTKVKDKFISDFYKKFRDKITVOGKNRKIALIPEVNEKVLKSEEGYFLLHLDLK
              FRIQPFETLQTLLERNDFNKRIRVKPIGIDFVGRVQDVFKAKEKGEFFRLCMERST
              HKSSKAWEELLKNRELREKAPIVVLEKGYTYPATILKPVLTYENLEDEERNEVADIV
              RMEPGKRLNLIRYILRRYVKALRDYGWYISPEEERAKGKLNFKDTVLDAKGKNTKVIT
              NLRKFLELCRPFVKKDVLSVEIISVSVYKKLLEWRKEEFLKELINFLKNKGIKKLIKGK
              SLILAQTREEAKEKLIPVINKIKDVDLVIVFLEEYPKVDPYKSFLLYDFVKRELLKKM
              IPSQVILNRTLKNENLKFVLLNVAEQVLAKTGNIPYKLKEIEGKVDAFVGIDISRITR
              DGKTVNAVAFTKIFNSKGELVRYLITSYPAFGEKLTEKAIGDVFSLLEKLGFKKGSKI
              VVHRDGRLYRDEVAAFKKYGELYGYSLELLEIIKRNNPRFFSNEKFIKGYFYKLSEDS
              VILATYNQVYEGTHQPIKVRKVYGELPVEVLCSQILSLTLMNYSSFQPIKLPATVHYS
              DKITKLMLRGIEPIKKEGDIMYWL"
gene          complement(2733..3152)
              /gene="mutT"
              /note="aq_1449"
CDS           complement(2733..3152)
              /gene="mutT"
              /note="aq_1449"
              /codon_start=1
              /transl_table=11
              /product="8-OXO-dGTPase domain (mutT domain)"
              /protein_id="AAC07397.1"
              /db_xref="GI:2983843"
              /translation="MGFNVKTPLLATDVIIRLWDGENFKGIVLIERKYPPVGLALPGG
              FVEVGERVEEAAAREMREETGLEVRLHKLMGVYSDPERDPRAHVVSVWIGDAQGEPK
              AGSDAKVKVYRLEEIPLDKLVFDHKKIILDFLKGNY"
gene          3181..4542
              /gene="htrA"
              /note="aq_1450"
CDS           3181..4542
              /gene="htrA"
              /codon_start=1
              /transl_table=11
              /product="periplasmic serine protease"
              /protein_id="AAC07399.1"
              /db_xref="GI:2983845"
              /translation="MRKFLLPLLIISVTVFSSTIKDLEEDIVKLVNKVAPSVVTIFVI
              KEEKSIFMPEFPFPDFPFPREFPTRKERSLGSGVIVKDEDKKVVVILTNAHVVKNG
              VRIIVKLDRHTEKKGEIVGIDTKTDIAVVKISTRGINDIEDRIAKLGDSDNLKVGQIV
              FAIGNPYEGLERTVMGVISALRRSIGITQYESFIQTDAAINPGNSGGPLINVEGEVIG
              INTAIIAGAQGLGFAIPINLAKWVMEQIIEHGKVIRGWLGVVIQDITPDISEALGIKE
              GVLVAQVVPGSPADKAGLKVGDVIVEVNGKKIEDARDLQFTIMKMKPGTKAVLKVIRN
              GKEKEITVIIGQYPEEGVSREGKATPENGLLLRDLTLHEKQEAGVPVGVLVGIIYPD
              SPAEYSGLQPGDIILKVNNRPVRSVREFYEIINRLKEMGRSKALLLVRRGDRNIFITL
              DLE"
gene          4553..5485
              /gene="rpoS"
              /note="aq_1452"

CDS           4553..5485
              /gene="rpoS"
              /codon_start=1
              /transl_table=11
              /product="RNA polymerase sigma factor RpoS"
              /protein_id="AAC07398.1"
              /db_xref="GI:2983844"
              /translation="MIPDTEQELINQYLKKISKIPLLTPEEKELARRAKQGDKEALK
              KLVESNLRFVVSVAKQYIGYGLPFSELIAAGNLGLLEAAKRFDPDRGVKFISYAVWWI
              RQAIMQALSQQTGAVRIPLKQSHIINKISHIYGELSSKKLEREPSPEIAEELSKQMIA
              KELEKELGREPTEEIEERFKKEGYKITPEEVEKYLQVCRVPLSLDAPVGENEDTFFI
              DFLSKHGTAEVEEKVIQEVIQKEIDDLLSKLTDKERKVIELRFGLKGNEPKTLREIGE
              ILGLSRERVRQLETRALRKLRSLAMKRHLKDFLS"
gene          5485..6297
              /gene="rpoS"
CDS           5485..6297
              /gene="aq_1453"
              /codon_start=1
              /transl_table=11
              /product="putative protein"
              /protein_id="AAC07403.1"
              /db_xref="GI:2983849"
              /translation="MKCVVYGSKSPVLHQALEEVKRKFLEEKILTDFDFMIISLNYKY
              PYENLHRDLVKIFDISSDAYFGFHSTEEIEETEIHDGLAVCFIKFENRGRLKVYWDSG
              ISDYMSNGLLQRLVEYLEENKNNLNVFFSAWEDKNLGLFIEDRILGOKVFNGEILRILG
              GVSSGRKFNGELRTFQFYKGKVIKDGFGILIFENVEFTLGISLGFKPISPIYEVKKAD
              NYKVYEVDSRVPFKRIVENFLTGLEKKIEYLWYCPIVLVEERRGIRKGAENL"
gene          6646..7356
              /gene="aq_1455"
CDS           6646..7356
              /gene="aq_1455"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="AAC07407.1"
              /db_xref="GI:2983853"
              /translation="MREDVKLRLLMNIIDEVTRKYNELLHNYKVLLKETEFEATHDPL
              TGLFNRTALQQLFKEKQRADVLGKKVAFLFIDLDNFKEVNDRYGHAEGDRVLKDIAR
              IIRESIRGTDIAVRIGGDEFLILPNSNIETAKRVGERIKNAIEETFKDYGISASYGI
              SIYPNEGETELEELINLADSRMYQMKREKKLKTGRYLKGNVHKGIFLLLLHYHCVNSET
              VLSYFNNSSNSYTCVLQN"
gene          complement(7213..7845)
              /gene="aq_1457"
CDS           complement(7213..7845)
              /gene="aq_1457"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="AAC07408.1"
              /db_xref="GI:2983854"
              /translation="MAHKFDPSKIKKLDDPSRLELFDPEKVLKEFGLKEGMTVLDVGT
              GAGFYLPYLSKMVGEKGKVYAIDVQEEMVNYAWEKVNKLGLKNVEVLKSEENKIPLPD
              NTVDFIFMAFTFHELSEPLKFLEELKRVAKPFAYLAIIDWKKEERDKGPPPEEVVSEW
              EVGLILEDAGIRVGRVVEVGKYCFGVYAMVKQEEENPLMNVPFKIPPGL"
gene          7990..8895
              /gene="gcvT"
              /note="aq_1458"
CDS           7990..8895
              /gene="gcvT"
              /codon_start=1
              /transl_table=11
              /product="aminomethyltransferase (glycine cleavage system
              T protein)"
              /protein_id="AAC07401.1"
              /db_xref="GI:2983847"
              /translation="MFDISHMGRLLIEDPEKKLQYFTTNNLDKLSVGKVQYNLLPNEK
              GGIRDDVTYMLSEFFLCVNAANRQKVINWLSPHLKLRDLSGELVQIALQGPKSEE
              IISKFYPVSDLKYYRFKVFDKTIISRTGYTGEDGPEIYVSPEGKELFLEVKLAKPC
              GLGARDVLRIEAGLPLYGNELSEITPIEVNLEKFVDFSKEFIGKEAMLKKKVKKKLF
              GLELTEKGIPRKGYRVFKGDREIGWISSGTYSPTLNKGIALCFVDIEERKEGNEVEIE
              VRGRRVRGVLRKYPFVRTPAGRYQK"
gene          complement(8886..10634)
              /gene="npr"
```

```
CDS          /note="aq_1459"
             complement(8886..10634)
             /gene="npr"
             /codon_start=1
             /transl_table=11
             /product="neutral protease"
             /protein_id="AAC07400.1"
             /db_xref="GI:2983846"
             /translation="MTLPIPSSPRIKINLPCFQYVGTEMSVDIITSIWFMLTEFAT
             FMPKRDMCPKPASIPPAAPTVENTVPTTIPQINRRSSITSSSSTTVDNFVIKIYQVYR
             RMQTLTLEISTTKRSVLNLKRASLTFQQFKNLLYLCLLEYYKNTKDVKPFLSVSFLE
             KFVKGKENLPFENEKIKSYREKLKELWEKEIGSDTVKALARQYSKEVKASVLEKWKKGE
             RAFLPRPRRLKKVHKFTVPTNPNMLVDKRKLKRRRENSVVRLGISFGAVKVKIPQGI
             NIRDVKITWYEDTEVIFRISYEVPKAQKELLKEKWLSIDLGVRNLVSCISNDENLRSF
             IVDGNPLKSFNQWVNKLSAKLRSEGKERLERKLWRYRKKRIKGFFHQVANLIVRVCLE
             RGIGRVIIPNSLNEEYQKDSEKGARFNQEFRFIPLGELVKEIMKIYKCEL/CGIEVVKEDE
             SWTSKVSSVSGNIEVLKEEIPEGEVKLKFEGRRIKRGLFRDRRLKKVFNADLNGALN
             LAIKALGKKVREEFLSLRNWLDKLSRPVKVNLFRYPASLPVLWEIAGSKSCSPPKGDSE
             GHFAGSVSDLITFAHF"
gene         10348..11238
             /gene="aq_1462"
CDS          10348..11238
             /gene="aq_1462"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAC07409.1"
```

```
alignment_scores:
     Quality:       259.00          Length:      188
     Ratio:         2.214           Gaps:        3
Percent Similarity:  62.234         Percent Identity:  34.574

alignment_block:
US-09-252-691-7056 x AE000741

Align seg 1/1  to: AE000741  from: 1  to: 15077

     37 GluProGlnProThrArgValIleLeuPheAsnLysProTyrAspValIle  53
        ::||||||    |||  ||    ||| ::|||||||||::||||||||| ||
  11409 AAGCCCCAGAGAAAACGGTACATAATCCTGAACAAGCCCTGTTGCTATT  11458

     53 uProGlnPhe...ThrAspGluAlaGlyArgSerThrLeuLysAspPheI  69
        |:::|||      ||| :|:|||::|||::||::|::|::::|||::||:|
  11459 GACTCAGCTCGGGAGGTeCCTGACGGGAGGAGACATTGAAGAGCTTA  11508

     69 lePro.......ValGlnGlyValTyrAlaAlaGlyAlaAspAlaAspAsp  83
        ||          ::: |||||||| :::||||||||||||||  :::
  11509 TAAAGGACATACCCGAGGGGTATTCCTGTGGGAAGGCTTGATTACAAC  11558

     84 SerGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLe  100
        ::|||||||||||||    :::|||||||||||||::||||||||||||||
  11559 ACTGAGGGCTTTTGCATACTAACCAACGACGGAGCTCCGCAAACAGGAT  11608

    100 uThrGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluG  117
        : :|||||     ||||:|:|||::|||::||::||||||||||||||||
  11609 ACTCCACCCGAGGTACAAACTGCCGAAGGTTTACCTCGCACTCGTTGAAG  11658

    117 lyGluProAspAspGluAlaSerLeuAlaLysLeuArgAsnGlyValThrLeu  133
        ||:::  |||::|: ::||: || |||||||||||:::||  |||
  11659 GGAAAGTAGACCAGAAAACTTTAAGAGAGGATGAAGCAGGAATAGAACTG  11708

    134 AsnAspGlyProThrLeuProAlaGlyLyIleGluArgValAsnGluProGl  150
        ::|||||||||||:  ||:||| ||||||||::||:::::|||||||||||
  11709 GAGGATGGCTTTGCAAAACCCGACAACATAAGGATAGTA..........  11747

    150 uTrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProT  167
        ||  :::||||| :|:|||| |||:||   ||:||  ::|||||||||
  11748 ...........CGCTACGAAGGGAAAA  11763

    167 hrSerTrpLeuLysIleThrLeuTyrGluGlyAsnArgGlnValArg  183
        :::|||||||||::||||||| |||||||::|||||||||||   |||||
  11764 ACACACTCCTTGAGATAACATTCCACGAAGGGAGGAAGCACTTAGTTAAG  11813
```

```
         184 ArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAl  200
             |||::|||    |:: :::    ||| ||||||  ||||||  |||||| ||
       11814 AGGTTTTAGGAGCTTCGGACATAAGGTAAAGCGTTTAAAGAGGATAGC  11863

         200 aMetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAspV  217
             |:::::||   ||    |:::|||     |||::|||||||::|::||||:
       11864 TATCGGACCAATAAAACTCGTAAGCTCTCTCCGGCAAGTGGAGGGAAC  11913

         217 alThrProLysGlu 221
             ::::: |||||||
       11914 TTAATCAGGGGGAA 11927
```

```
seq_name: gb_bal:BACDIA

seq_documentation_block:
LOCUS          BACDIA         28206 bp    DNA            BCT        26-MAY-1995
DEFINITION     Bacillus subtilis spoVA to serA region.
ACCESSION      L09228
VERSION        L09228.1  GI:410114
KEYWORDS       3-dehydroquinate dehydratase; aroC gene; diaminopimelate
               decarboxylase; lysA gene; penicillin-binding protein;
               peptidyl-prolyl isomerase; phosphoglycerate dehydrogenase; ppiB
               gene; response regulator; response regulator kinase; ribA gene;
               ribB gene; ribD gene; ribG gene; ribH gene; ribT gene; riboflavin
               biosynthesis operon; serA gene; signal peptidase; sipS gene; spoA
               gene; spoVAF gene.
SOURCE         Bacillus subtilis (strain 168, sub_species Marburg) DNA.
ORGANISM       Bacillus subtilis
               Bacteria; Firmicutes; Bacillus/Clostridium group;
               Bacillus/Staphylococcus group: Bacillus.
REFERENCE      1  (bases 1 to 28206)
AUTHORS        Yamamoto,J., Shimizu,M. and Yamane,K.
TITLE          Molecular cloning and analysis of nucleotide sequence of the
               Bacillus subtilis lysA gene region using B. subtilis phage vectors
               and a multi-copy plasmid, pUB110
JOURNAL        Agric. Biol. Chem. 55 (6), 1615-1626 (1991)
MEDLINE        91345841
REFERENCE      2  (bases 1 to 28206)
AUTHORS        Buchanan,C.E. and Ling,M.L.
TITLE          Isolation and sequence analysis of dacB, which encodes a
               sporulation-specific penicillin-binding protein in Bacillus
               subtilis
JOURNAL        J. Bacteriol. 174 (6), 1717-1725 (1992)
MEDLINE        92193254
REFERENCE      3  (bases 1 to 28206)
AUTHORS        Sorokin,A., Zumstein,E., Azevedo,V., Ehrlich,S.D. and Serror,P.
TITLE          The organization of the Bacillus subtilis 168 chromosome region
               between the spoVA and serA genetic loci, based on sequence data
JOURNAL        Mol. Microbiol. 10 (2), 385-395 (1993)
MEDLINE        95020538
FEATURES             Location/Qualifiers
     source          1..28206
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /sub_species="Marburg"
                     /db_xref="taxon:1423"
     gene            1..1239
                     /gene="spoVAF"
     CDS             <1..1239
                     /gene="spoVAF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA67472.1"
                     /db_xref="GI:410115"
                     /translation="VDIVENRLLNAQVEKVKTLDETTDQVLSGLVAVIVEGAGFAFII
                     DVRSYPGRNPEPDTEKVVRGARDGFVENIVVNTALLRRRIRDERLRVKMTKVGERSK
                     TDLSICYIEDIADPDLVEIVEKEIASIDVDGLTMADKTVEFIVNQSYNPFPLVRYTE
                     RPDVRAANHVLEGHVIIVDTSPSVIITPTLFHHVQHAEEYRQAPSVGTFLRWRFFG
                     ILASTLFLPIWFLVFVLQPDLLPDNMKFIGLNKDTHIPIILQIFILADLGIEFLRMAAIH
                     TPTALSTAMGLIAAVLIGQIAIEVGLFSPEVILYVSLAAIGTFTTPSYELSLATNEPS
                     CPHDTRCFISYKRARHRLYSANYAMASIKSLQTPYLWPLIPFNGKALWQVLVRTAKPG
                     AKVRPSIVHPKNRLROPTNS"
```

```
conflict        1158..1164
                /gene="spoVAF"
                /citation=[1]
-35_signal      1245..1250
                /gene="lysA"
                /note="putative"
gene            1245..2664
                /gene="lysA"
-10_signal      1266..1272
                /gene="lysA"
                /note="putative"
CDS             1345..2664
                /gene="lysA"
                /codon_start=1
                /transl_table=11
                /product="diaminopimelate decarboxylase"
                /protein_id="AAA67473.1"
                /db_xref="GI:410117"
                /translation="MFLHGTSRQNQHGHLEIGGVDALYLAEKYGTPLYVYDVALIRER
                AKSFKQAFISAGLKAQVYASKAFSSVAMIQLAEEEGLSLDVVSGGELYTAVAAGFPA
                ERIHFHGNNKSREELRMALEHRIGCIVVDNFYEIALLEDLCKETGHSIDVLRITPGV
                EAHTHDYITTGGEDSKFGFDLNGQTERAIEQVLQSEHIQLLGVHCHIGSQIFDTAGF
                VLAAEKIFKKLDEWRDSYSFVSKVLNLGGGFGIRYTEDDEPLHATEYVEKIIEAVKEN
                ASRYGFDIPEIWIEPGRSLVGDAGTTLYTVGSQKEVPGVRQYVAVDGGMNDNIRPALY
                QAKYEAAAANRIGEAHDKTVSIAGKCCESGDMLIWDIDLPEVKEGDLLAVFCTGAYGY
                SMANNYNRIPRPAVVFVENGEAHLVVKRETYEDIVKLDLPFKTGVKQ"
conflict        1610..1616
                /gene="lysA"
                /citation=[1]
conflict        1751..1757
                /gene="lysA"
                /citation=[1]
conflict        2496..2502
                /gene="lysA"
                /citation=[1]
conflict        2504..2510
                /gene="lysA"
                /citation=[1]
terminator      2666..2694
                /gene="lysA"
terminator      complement(2666..2694)
                /note="ORFX19"
CDS             complement(2693..3565)
                /note="ORFX19"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA67474.1"
                /db_xref="GI:410118"
                /translation="MKKIWIGMLAAAVLLMVPKVSLADAAVGDVIVTLGADLSESDK
                QKVLDEMVPDNATTVTVTNKEEHEYLGKYISNAQIGSRAISSSSIIAKKGSGLNVE
                THNISGITDEMYLNALMTAGVKDAKVVTAPFEVSGTALTGLIKAYEVSSDEAISED
                VKQVANQELVTTSELGDKIGNENAAALIAKIKEEFAKNGVPDNKADIEKQVDDAASDL
                NVTLTDSQKNQLVSLFNKMKNADIDWGQVSDQLDKAKDKIITKFIESDEGKNFIQKVID
                FFVSIWNAIVSIFK"
conflict        3347..3353
                /citation=[1]
conflict        3413..3419
                /citation=[1]
conflict        3428..3434
                /citation=[1]
conflict        3503..3509
                /citation=[1]
gene            3819..4290
                /gene="ppiB"
CDS             3819..4250
                /gene="ppiB"
                /codon_start=1
                /transl_table=11
                /product="peptidyl-prolyl isomerase"
                /protein_id="AAA67475.1"
                /db_xref="GI:410119"
                /translation="MKTGYFLLEDGNKIEFELYPEAAPGTVANFEKLANEGFYDGLTF
                HRVIPGFVSQGGCPHGTGTGGPGYTIKCETEGNPHTHEAGALSMAHAGKDTGGSQFFI
                VHEPQPHLNGVHTVFGKVTSGLEFAKNMSNGDVMKEVRVEG"

conflict        3982..3988
                /gene="ppiB"
                /citation=[1]
terminator      4261..4290
                /gene="ppiB"
repeat_region   4287..4320
                /note="inverted repeat"
CDS             5264..5467
                /note="ORFX1"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA67476.1"
                /db_xref="GI:410120"
                /translation="MQVLTGLLEQYGYAVLFIVMLELLALLGPEFAGVLVFQEKLN
                WVLSMLAAGLGSCGSEHVVLDRV"
CDS             5445..5834
                /note="ORFX2"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA67477.1"
                /db_xref="GI:410121"
                /translation="MSYNIGYRLGPAAIQKAFASRGVLLAVIVLIFFWTVFRGQDRKI
                ELSFLFFLVIGGELYEEGIRRLFHSLHPVHAPLESHVLYPFPSEQTLTAFVIFGFCAY
                LFVRHSHKTRLQTVAFFCLLSFYISSSG"
-35_signal      6022..6027
                /gene="sipS"
                /note="putative"
gene            6022..7323
                /gene="sipS"
-10_signal      6045..6051
                /gene="sipS"
                /note="putative"
-35_signal      6683..6688
                /gene="sipS"
                /note="putative"
-10_signal      6705..6710
                /gene="sipS"
                /note="putative"
CDS             6740..7294
                /gene="sipS"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="signal peptidase"
                /protein_id="AAA67478.1"
                /db_xref="GI:410122"
                /translation="MKSENVSKKKSILEWAKAIVIAVVLALLIRNFIFAPYVVDGDSM
                YPTLHNRERVFVNMTVKYIGEFDRGDIVLNGDDVHYVKRIIGLPGDTVEMKNDQLYI
                NGKKVDEPYLAANKKRAKQDGFDHLTDDFGPVKVPDNKYFVMGDNRRNSMDSRNGLGL
                FTKKQIAGTSKFVYPFNEMRKTN"
terminator      7299..7323
                /gene="sipS"
-35_signal      7388..7393
                /gene="sipS"
-10_signal      7410..7416
                /note="putative; ORFX4"
CDS             7529..7873
                /note="ORFX4"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA67479.1"
                /db_xref="GI:410123"
                /translation="MGRIKTKITLLVLLLLAGGYMYINDIELKDVPTAIGQTLSSE
                EEEYTIQEYKVTKIDGSEYHGVAENGTKIIFNGKKLNQDLSDIKEGDKIKAYFSKSKR
                IDGLIKVAKVND"
terminator      7872..7911
                /note="ORFX4"
-35_signal      7937..7942
-10_signal      7960..7965
protein_bind    8126..8137
                /note="putative"

alignment_scores:
               Quality:   253.00          Length:    182
```

```
Ratio:            2.259              Gaps:              5
Percent Similarity:  61.538    Percent Identity:  35.165

alignment_block:
US-09-252-691-7056 x BACDIA

Align seg 1/1  to: BACDIA  from: 1  to: 28206

  39 GlnProThrArgValIleLeuPheAsnLysProTyrAspValLeuProGl 55
     ::||||:::      :::||||::      |||||      |||||
17614 GAGCCGGTTACTTCTCTTATAT...AAACCAAGAGGCGTTATTCCGC 17660

  55 nPheThrAspGluAlaGlyAlaGlyThrLeuLysAspPhe.........I 69
     |||||          |||||||          ||||||
17661 AGCGCAAGATGATAAAGA ACGCAAAGTGGTGACGGACTTTTTAAAAATA 17710

  69 leProValGlnGlyValTyrAlaAlaGlyAlaArgLeuAspArgAspSerGlu 85
     |||||      |||      ::||||||      |||::::||      ::|||
17711 TTCCG...CAGCGCAATTATCCGATTGGGCGTCTCGGACTATGATACGAGC 17757

  86 GlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGl 102
     |||||||||          ::||      ::|||      :::|||      ::
17758 GGTCTTTTGCTTTAACCAATGACGGCGAATTCGCCAATAAGCTAATGCA 17807

 102 nProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGluP 119
     :|||      |||      |||||      :::|||::|||
17808 TCCTAAGTATGAAATAGACAAAACATACGTGGCGAAGGTGAAAGGAATCC 17857

 119 roAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAsp 135
     ||          |||||::::||||:::||||::      |||::||||
17858 CGCCTAAAGAGCTGCTCAGAAAGCTGGAGCCGCGGAATTAGGCTGGAGGAA 17907

 136 GlyProThrLeuProAlaGlyIleGluArgValAsnGluProGluTrpLe 152
     |||      |||          :::|||      |||||      |||||
17908 GGCAAAACAGCGCCTGCTAAAGCAAGCGAAGCTGCTGTCT........... 17943

 152 uTrpProArgAsnProProIleArgLysSerIleProThrSerT 169
     :::      ::||||:::      |||||      ||||||
17944 ..................TTGGATAAGAAAAAGCAG......ACAAGCA 17968

 169 rpLeuLysIleThrLeuTyrGluGlyTyrArgAsnArgGlnValArgArgMet 185
     ::::||||::||||:::      ||||||||||||||
17969 TTATTCAGCTGACGATCCCATGAGGGACGCAACAGAAGTGCGCCGCATG 18018

 186 ThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGl 202
     :::||||      ||||||      |||||      |||||||
18019 TTTGAAGGCGATCGGACATGAGGTAATCAAGCTGAAGCGTGAGGAATACGC 18068

 202 ySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAspVal 217
     ::|||      :::||||:::      |||::::||
18069 GTTTTAAACCTGCGAGGCCTGCACACGGGCGATGCCCAGAGAGTTA 18114

seq_name: gb_bal:BSUB0013
```

```
seq_documentation_block:
LOCUS       BSUB0013   218470 bp   DNA                 BCT        26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 13 of 21): from 2395261
            to 2613730.
ACCESSION   299116 AL009126
VERSION     299116.1  GI:2634723
KEYWORDS    .
SOURCE      Bacillus subtilis.
ORGANISM    Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1  (bases 1 to 218470)
AUTHORS     Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
TITLE       The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
JOURNAL     Nature 390 (6657), 249-256 (1997)
MEDLINE     98044033
REFERENCE   2  (bases 1 to 218470)
AUTHORS     Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1..218470
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     terminator      complement(7..18)
                     /gene="cmk"
     gene            complement(7..696)
                     /gene="cmk"
     CDS             complement(22..696)
                     /gene="cmk"
                     /EC_number="2.7.4.14"
                     /note="alternate gene name: jofC, ypfC"
                     /codon_start=1
                     /transl_table=11
                     /product="cytidylate kinase"
                     /protein_id="CAB14221.1"
                     /db_xref="GI:2634724"
                     /db_xref="SWISS-PROT:P38493"
                     /translation="MEKKLSIAIDGPAAAGKSTVAKIVAEKKSYIYIDDGAMYRAITY
                     AALQENVDLTDEEKLAELKRTDIELITYKDGQKVFVNGTDVTEAIRTDEISNQVSIA
                     AKHRSVREEMVKRQQOLGEKGGVVMDGRDIGTHVLPAAEVKIFLLASVERAKRRYEE
                     NVKKGFDVNYETLIIEIARRDKLDSEREVSPLRKAEDALEIDTTSLSIQEVADKILEA
                     VEQKSR"
     gene            complement(775..951)
                     /gene="ypfB"
     CDS             complement(775..951)
                     /gene="ypfB"
                     /function="unknown"
                     /note="alternate gene name: jofB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB14222.1"
                     /db_xref="GI:2634725"
                     /db_xref="SWISS-PROT:P38492"
                     /translation="MKTFERLLIKLLFIQAILLGVQFLFHYQHIEPYVSKVIQYEGV
                     DKMEENRIETFKH"
     gene            complement(996..1649)
```

CDS
```
/gene="ypfA"
complement(996. .1649)
/gene="ypfA"
/function="unknown"
/note="alternate gene name: jofA"
/codon_start=1
/transl_table=11
/protein_id="CAB14223.1"
/db_xref="GI:2634726"
/db_xref="SWISS-PROT:P38491"
/translation="MIEIGENVLLEYIEENELKKAKSKAVSIENNELLIAYPVDVVTG
RTVILHNDMEVTVEFVGKDEVPYRFISRIKGKVKDKLQMICLEMPPREKMKRIQRRQY
VRTDAVLDVQIOPGNEEEIRTLSRYNISAGGIAVVLADGLSFQSGESLRLIIRLPEEEH
TRQIETEAVVRRIFNDPKSEKRKMTLEYSEIAAGDQQALLQYCIRRQLNKKRKARME"
```

gene
```
/gene="ypeB"
complement(1711. .3094)
/gene="ypeB"
```

terminator
```
complement(1711. .1733)
/gene="ypeB"
```

CDS
```
complement(1742. .3094)
/gene="ypeB"
/function="unknown"
/note="alternate gene name: joeB, yzuA; similar to
hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14224.1"
/db_xref="GI:2634727"
/db_xref="SWISS-PROT:P38490"
/translation="MIRGILIAVLGIRAIVGTGWGYKEHQEKDAVLLHAENNYQRAFH
ELTYQVDQLHDKIGTTLAMNSQKSLSPALIDVWRITSEAHNSVSQLPLTLMPFNKTEE
LLSKIGDFSYKTSVRDLDQKPLDKNEYTSLNKLYQOSEDIONELRHVQHLVMSKNLRW
MDVEMALASDEKQSDNTIINSFKTVEKNVGAFSTGTDLGPSFTSTKKEKGFSHLKGK
QISEQEAKQIAERFAPDDNYSIKVVVKSGGKKTNRDVYSISMKDPDHKAVIYMDITKGG
HPVYLIQNREKVDQKISLNDGSNRALAFLKKNGFETDDLEIDESAQYDKIGVFSYVPV
ENKVRMYPEAIRMKVALDDGEVVGFSARDFLTSHRKRTIPKPAITEEAKSKLNKNVQ
VRETRLALITNELGOEVLCYEMLGTIENDTFRMYINAKDGSEEKVEKLKNAEPIYKDL
"
```

gene
```
complement(3129. .4046)
/gene="sleB"
```

CDS
```
complement(3129. .4046)
/gene="sleB"
/note="alternate gene name: ypeA"
/codon_start=1
/transl_table=11
/product="spore cortex-lytic enzyme"
/protein_id="CAB14225.1"
/db_xref="GI:2634728"
/db_xref="SWISS-PROT:P50739"
/translation="MKSKGSIMACLLILFSFTITTFINTETISAFSNQVIQRGATGDDV
VELQARLQYNGYYNGKIDGVYGWGTIWAVRNFQDQFGLKEVDGLVGAKTKQTIICKSK
YYREYVMEQLNKGNTFTHYGKIPLKYQTKPSKAATQKARQQAEEAROKQPAEKTTQKPK
ANANKQQNNTPAKARKODAVAANMPGGFSNNDIRLLAQAVYGEARGEPYEGQVAIAAV
ILNRLNSPLFPNSVAGVIFEPLAFTAVADGQIIMQPNETAREAVLDRAINGWDPSEEAL
YFNPDTATSPWIWGRPQIKRIGKHIFCE"
```

gene
```
complement(4162. .4841)
/gene="ypdC"
```

terminator
```
complement(4162. .4182)
/gene="ypdC"
```

CDS
```
complement(4185. .4841)
/gene="ypdC"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14226.1"
/db_xref="GI:2634729"
/db_xref="SWISS-PROT:P50738"
/translation="MFAIISAGIAPGIALLSYFYLKDQYDNEPVHMVLRSFFLGVVLV
FPIMFIOYVLEKENVGGGSFFVSFLSSGFLEESLKWFILMISVYPHAHFDEHYDGIVY
GASVSLGFATLENILYLIGHGVEHAFVRALLFVSCHALIGVIMGFYLGKARFSADKAR
VKWLTLSLVVPSLLHGSYDFILTALSNWIYMLPFMVFLWWFGLRKAKKARSVNMMQV
"
```

gene
```
complement(4929. .5935)
```

terminator
```
/gene="ypdA"
complement(4929. .4953)
/gene="ypdA"
complement(4961. .5935)
/gene="ypdA"
/function="unknown"
/note="similar to thioredoxin reductase"
/codon_start=1
/transl_table=11
/protein_id="CAB14227.1"
/db_xref="GI:2634730"
/db_xref="SWISS-PROT:P50736"
/translation="MIQGKAIIIGGPCGLSAAIHLKQIGIDALVIEKGNVVNSIYNY
PTHQFFSSSEKLEIGDVAFITENRKPVRIQALSYYREVKRKNIRVNAFEMVRKVTK
TQNNTFVIETSKETYTTPYCIIATGYYDHPNYMGVPGEDLPKVFHYFKEGHPYFDKDV
VVLGGKNSSVDAALELVKSGARVTVLYRGNEYSPSIKPWILPEFEALVRNGTIRMEFG
ACVEKITENEVVFRSGEKELITIKNDFVFAMTGYHPDHQFLEKIGVEIDKETGRPFFN
EETMETNVEGVFIAGVIAAGNNANEIFIENGRFHGGHIAAEIAKRENH"
```

gene
```
/gene="ypcA"
complement(5995. .7324)
```

terminator
```
complement(5995. .6016)
/gene="ypcA"
```

CDS
```
complement(6044. .7324)
/gene="ypcA"
/function="unknown"
/note="similar to glutamate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAB14228.1"
/db_xref="GI:2634731"
/db_xref="SWISS-PROT:P50735"
/translation="MAADRNTGHTEEDKLDVLKSTQTVIHKALEKLGYPEEVYELLKE
PMRLLTVKIPVRMDDGSVKIFTGYRAHNDSVGPTKGGIRFHPNVTEKEVKAVKALSIW
MSLKCGIIDLPYGGGKGGIVCDPRDMSFRELERLSRGVRAISQIVGPPHKDVPAPDVE
TNSOIMAWWMDEYSRIDEFNSPGFITGKPLVLGGSHGRESATAKGVTICIKEAAKKRG
IDIKGARVVVQGFGNAGSYLAKFMHDAGAKVVGISDAYGGLYDPEGLDYDYLLDRRDS"
```

```
alignment_scores:
    Quality:        253.00          Length:          182
    Ratio:            2.259          Gaps:              5
Percent Similarity:  61.538    Percent Identity:   35.165

alignment_block:
US-09-252-691-7056 x BSUB0013/rev

Align seg 1/1  to reverse of: BSUB0013  from: 1  to: 218470

   39 GlnProThrArgValIleLeuPheAsnLysProTyrAspValLeuProGl 55
      ||||| || ||||  :: ||| ::|:|: ||||       ||||||
25969 GAGCCGGTTTACTTTCTCTTATAT...AAACCAAGAGGCCGTTATTTCCGC 25923

   55 nPheThrAspGluAlaGlyArgSerThrLeuLysAspPhe.........I 69
      |||:: ||||||::   |||||||||||   :||||        ||
25922 AGCGCAAGATGATAAAGGACGCAAAGTGGTGACGGACCTTTTTAAAAATA 25873

   69 leProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGlu 85
      ||||| ||| ||| ::::|||  |||||||||||||| |||:::::|:
25872 TTCCG...CAGCGCATTATCCGATTGGGCGTCTGGACTATGATACGAGC 25826

   86 GlyLeuLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGl 102
      |||||| || || ||||||||||||| :: |||||||   ::|||
25825 GGTCTTTTGCTTTAACCAATGACGGCGAATTCGCCAATAAGCTAATGCA 25776

  102 nProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGluP 119
      :||| ::: |||||||||||:||| :::||:||||||||||||||| ::
25775 TCCTAAGTATGAAATAGACAAAACATACGTGGCGAAGGTGAAAGGAATCC 25726

  119 roAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAsp 135
      || || |||||||:|:|:|||||||| ||||||:||||||   |||||:
25725 CGCCTAAAGACGCTGCTCAGAAAGCTGAGCCGCCGGAATTAGGCTGGAGGAA 25676

  136 GlyProThrLeuProAlaGlyIleGluGlyArgValAlaAsnGluProGluTrpLe 152
      ||| ||| ||||       :: ::: ||||| :  :::::
```

```
25675 GGCAAAACAGCGCCTGCTAAAGCGAAGCTGCTGTCT............. 25640
  152 uTrpProArgAsnProProIleArgGluArgGlyLysSerIleProThrSerT 169
                         ::  :.:::::|||::::     ||||||
25639 ...............TTGGATAAGAAAAAGCAG......ACAAGCA 25615
  169 rpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMet 185
      ::::::::|||::|||||||||||||||||||||||
25614 TTATTCAGCTGACGATCCATGAGGGACGCAACAGACAAGTGCGCCGCATG 25565
  186 ThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGl 202
      ::|||      :::::::|||  |||      ::
25564 TTTGAAGCGATCGGACGACATGAGGTAATCAAGCTGAAGCCGTGAGGAATACGC 25515
  202 ySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAspVal 217
      :  :::||||  :::|||||  :::|||:::::
25514 GTTTTTAAACCTGCGAGGCCTGCACACGGGCGATGCCAGAGAGTTA 25469
```

OM of: US-09-252-691-7056 to: N_Geneseq_36:*   out_format : pfs

Date: Jun 10, 2000 1:19 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -O=/cgn2_1/USPTO.spool/US09252691/runat_05062000_101651_631/app_query.fasta.1
 -DB-N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -ALIGN-15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
 -THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 59.520000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X13433	+	283.50	506.38	1.3e-20	1075	! Enterococcus faecalis genome c
N_Geneseq_36:V52297	+	246.50	415.64	1.5e-15	9707	! Streptococcus pneumoniae genom
N_Geneseq_36:V65255	+	230.50	391.21	3.4e-14	5816	! DNA encoding S. pneumoniae prc
N_Geneseq_36:V52369	+	222.50	397.14	1.6e-14	748	! Streptococcus pneumoniae genome
N_Geneseq_36:X13358	-	208.00	359.01	2.1e-12	2285	! Enterococcus faecalis genome c
N_Geneseq_36:V22799	+	206.00	352.35	5.0e-12	3061	! Nucleotide sequence of the E3
N_Geneseq_36:V74482	+	202.50	332.82	6.2e-11	11126	! Staphylococcus aureus contig
N_Geneseq_36:X13710	-	202.00	355.47	3.4e-12	1081	! Enterococcus faecalis genome c
N_Geneseq_36:V74878	-	201.50	359.30	2.1e-12	676	! Staphylococcus aureus contig SE
N_Geneseq_36:V52353	+	185.50	310.09	1.1e-09	4692	! Streptococcus pneumoniae genom
N_Geneseq_36:V52508	+	167.00	294.80	1.1e-09	720	! Streptococcus pneumoniae genome
N_Geneseq_36:X20248_07	+	151.00	214.32	0.0002	110000	! Continuation (8 of 10) of
N_Geneseq_36:T84021	-	128.50	225.66	5.7e-05	584	! DNA encoding a Staphylococcus a
N_Geneseq_36:V65214	+	127.00	206.09	0.0007	3071	! DNA encoding a S. pneumoniae p
N_Geneseq_36:V52139	+	126.00	185.19	0.0103	20199	! Streptococcus pneumoniae genc
N_Geneseq_36:V20516	-	119.50	181.74	0.0160	8642	! Polynucleotide sequence from t
N_Geneseq_36:V52155	+	99.00	131.85	9.62	28171	! Streptococcus pneumoniae genc
N_Geneseq_36:V22674	-	94.50	117.65	59.46	50341	! DNA sequence of a L5 shuttle
N_Geneseq_36:T51411	-	94.50	117.26	62.48	52297	! Mycobacteriophage L5 genome s
N_Geneseq_36:Q47357	-	94.50	117.26	62.48	52298	! L5 mycobacteriophage DNA. Myc
N_Geneseq_36:X13111	-	89.50	135.37	6.12	3496	! Enterococcus faecalis genome c
N_Geneseq_36:X07134	+	88.50	147.14	1.35	909	! Staphylococcus aureus mutant P1
N_Geneseq_36:T04318	+	87.00	127.91	15.93	4624	! Expression vector pUC19SD5 fra
N_Geneseq_36:Q41259	-	86.00	131.74	9.76	2639	! 2,2-dialkylglycine decarboxyla
N_Geneseq_36:Q72718	+	86.00	131.74	9.76	2639	! Dialkylglycine decarboxylase g
N_Geneseq_36:Q62811	+	85.50	129.23	13.46	3086	! Riboflavin 2 gene (rib-2). DNA
N_Geneseq_36:T80413	+	85.50	102.52	413.98	43280	! Tylactone synthase gene clust
N_Geneseq_36:V32452	+	84.50	116.57	68.25	8982	! Human receptor tyrosine kinase
N_Geneseq_36:Q11707	+	84.50	111.43	131.93	14928	! Immunoglobulin D-region heavy
N_Geneseq_36:V31354	+	84.00	123.74	27.21	4034	! Trichoderma reesei ACEI transc
N_Geneseq_36:T33238	+	83.50	117.97	57.03	6513	! Drosophila para voltage-activa
N_Geneseq_36:T18986	-	83.50	117.97	57.03	6513	! Drosophila para voltage-activa
N_Geneseq_36:V65267	+	83.00	135.37	5.82	1023	! DNA encoding two S. pneumoniae
N_Geneseq_36:Q14412	+	81.00	112.34	117.49	7195	! Forward strand of Burmese ET-N
N_Geneseq_36:X20580	-	81.00	102.44	418.11	19142	! Polynucleotide sequence from
N_Geneseq_36:Q12401	+	80.50	132.06	9.36	934	! Gene encoding dihydrodipicolina
N_Geneseq_36:X19362	+	80.50	110.88	141.65	7584	! Rhodococcus corallina ohp oper
N_Geneseq_36:V21187	+	80.50	91.06	1.8e+03	53789	! Amycolatopsis mediterranei ri
N_Geneseq_36:T58840_2	-	80.50	83.83	4.6e+03	110000	! Continuation (3 of 6) of T5
N_Geneseq_36:X20228	+	80.00	112.04	121.99	6168	! Enterococcus faecalis gene EFl
N_Geneseq_36:X20230	+	80.00	112.04	121.99	6168	! Enterococcus faecalis gene EFl

N_Geneseq_36:X20232	+	80.00	112.04	121.99	6168	! Enterococcus faecalis gene
N_Geneseq_36:X13101	+	80.00	96.97	842.63	27360	! Enterococcus faecalis geno
N_Geneseq_36:V71205	-	79.50	123.29	28.82	1851	! DNA encoding esterase El01
N_Geneseq_36:T86704	-	79.50	117.92	57.38	3147	! DNA encoding thermostable e

seq_name: N_Geneseq_36:X13433

seq_documentation_block:
ID X13433 standard: DNA; 1075 BP.
AC X13433;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:496.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuate an Enterococcus
PT infection.
PS Claim 1; Page 1742-1743; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcus faecalis infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 1075 BP;   361 A;   169 C;   225 G;   318 T;

alignment_scores:
     Quality: 283.50   Length: 179
     Ratio: 2.423   Gaps: 3
     Percent Similarity: 65.363   Percent Identity: 34.637

alignment_block:
US-09-252-691-7056 x X13433

Align seg 1/1 to: X13433   from: 1 to: 1075

 45 LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGl   61
     :|||||| :|||||||||||||| ::||||||||:|||||
 76 TTATTCTATAAACCAGAGAGGGTTATTTCTCGGTGTCTGACGATAAAGG   125

 61 yArgSerThrLeuLysAspPheIlePro......ValGlnGlyValTyrA   76
     |||| :|||||||||||||| |||::|||||||||||||
126 CCGTAAAGTAGTAACGACGATTATTTTACAGATGTTACTGACGGAATTTATC   175

 76 laAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsn   92
     :|||||||||||| ||||:::::||||||||||||||:||||||||
176 CAGTCGGTCGGTCTCGGATTATGATACTTCTGGGTTGTTACTTTTAACGAAT   225

 93 AspGlyValLeuGlnIleAlaArgLeuThrGlnProGlyLysArgThrGlyLy   109
     :::||||||||||:::||||||||||::|||||||:|::||
226 GATGGAAGTTTTTCACAAAATTAACCATCCTAAACATGAAGTAGATAA   275

109 sIleTyrTyrValGlnValGluGlyGluGluProAspAspAlaSerLeuAlaL   126

```
alignment_scores:
        Quality:    246.50        Length:      179
        Ratio:        2.221        Gaps:          3
Percent Similarity:  62.011   Percent Identity:  33.520

alignment_block:
US-09-252-691-7056 x V52297

Align seg 1/1  to: V52297  from: 1  to: 9707

 45 LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGlyIl  61
    ||||:|||||||| :||::: |||::::    |||||||:: |
5010 CTGCTTAACAAACCACGCGGTGTGATTCCAGTGTGACAGATGATAAGGG  5059

 61 yArgSerThrLeuLysAspPheIlePro.....ValGlnGlyValTyrA  76
    |||::||::|||::||::||         |||:::  ||||||||
5060 TCGCAAGACGGTTGTCGACCTCTTGCCAATGTCAAAGAGCGTATTTACC  5109

 76 laAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsn  92
    :::|||:||||||||::|||||| |||:::|||||::: |||||||
5110 CTGTGGGTCGTTGGACTGGGATACATCAGGTGTCTGATTTGACCAAT  5159

 93 AspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLy  109
    ||||| ||  |||:||||:||||||: |||::: ||||||
5160 GATGGGGACTTTACAGCACGAGATGATTCACCCTCGTAATGAGATTGACAA  5209

109 sIleTyrTyrValGlnValGluGlyValGluProAspAspAlaSerLeuAlaL  126
    |:::|||   :|::|||::|||:||||        :::||||
5210 GGTTTATGTCCGCGGTGTTAAAGGTGGCCAATAAGGACAATCTCCGCC  5259

126 ysLeuArgAspGlyValThrLeuAsnAspGlyProThrLeuProAlaGly  142
    ||:|||        :::||||||||::|::::|||::
5260 CCTTGACCCGTGGTCTTGAGATTGAGGTAAGAAAACCAAGCCAGCTGTT  5309

143 IleGluArgValAsnGluProGluTrpProAsnProProIl  159
    |||                                ::: |||
5310 TATGAAATTCTC........................AAAGTGGACCCCAGT  5335

159 eArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrG  176
    :::|||||                :||:|||:::::|||:||::|
5336 CAAAAATCGC..........TCTGTGGTGCAGTTGACCATCCATG  5370

176 luGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhePro  192
    |||||||||:::|||||:::|||||||::: |||
5371 AAGGGCGTAACCATCAGGTTAAAAGATGTTGAAGCGTGTTGGTCTCCAA  5420

193 ThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSerLe  209
    ::: |||||||       ||| ||   |||||||:::     :::||
5421 GTAGATAAGTGTCTCGGACTCGTTTCGGACACCTAGACTTGACAGGACT  5470

209 uAlaAsnGlyGluTrpArgAspValThrProLysGlu  221
    |      ||||| :::|||  :::::   |||||
5471 CCGTCCAGGAGAATCCCGTCGTCTTAATAAAAAAGAA  5507
```

seq_name: N_Geneseq_36:V65255

(left column, continued)

```
 l::::|||    :::::||:::|||  |||                :::|||
276 AGTCTACGTAGCAAAAATTAAAGGTGTTCCAACTAAACGGGATCTATTGC  325

126 ysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGly  142
    |||||||:: |||||| |::::|     ||  |||||:||
326 CATTAGCCAAAGGCATTCGCATTGATGGCAAACGACCACCAGCCCAAT  375

143 IleGluArgValAsnGluProGluTrpProAsnProProIl  159
    ::::::||                          :::
376 TTTCAAATTCTT.......................TCAGCGGACAT  398

159 eArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrG  176
    |::    :::|||:::          :|||:::||:||:||:||::|
399 TAAACAGGTAGCAGTGTC.........GTGGAATTACGATTCATG  436

176 luGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhePro  192
    ||||||||||:::||||:::|||::|||:||::::   |||||
437 AAGGACGTAATCATCAAGTGAAGAAAATGTTCAAGCGTGTTCCCT  486

193 ThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSerLe  209
    ::: |||    :::||||  |||    |||||||:::       |||
487 GTTCAAAATTAAAACGTGAACGTTACGGTCAGTTGACCTTACAAGGTTT  536

209 uAlaAsnGlyGluTrpArgAspValThrProLysGlu  221
    |     :::||||:||:::|||  :::|||  |||||
537 GCGACCAGGGCAATATCGTGATTAACTAAAAAGAA  573
```

seq_name: N_Geneseq_36:V52297

PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI Young Bellido ML;
DR WPI: 98-348529/30.
DR P-PSDB; W80686, W80687, W80688, W80689, W80690, W80691, W80692.
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT for evaluating gene expression, and identification of virulence
PT genes
PS Claim 1; Pages 114-117; 333pp; English.
CC This sequence encodes Streptococcus pneumoniae proteins of unknown
CC functions. The invention provides DNA sequences (V65201 to V65304) from
CC the Streptococcus pneumoniae genome and corresponding protein sequences
CC (W80605 to W80728). A recombinant host containing a vector comprising any
CC of the above nucleic acids can be used for the recombinant expression of
CC the protein sequences. The invention also provides a DNA chip having
CC arrayed on it at least 15 base pair fragment of any one or more of these
CC DNA sequences. The DNA chip can be used methods for evaluating gene
CC expression in S. pneumoniae and for identifying virulence genes in
CC S. pneumoniae. Antibodies that selectively bind to the above proteins or
CC peptide fragments can be used to treat S. pneumoniae infection. The
CC antibodies can also be used to detect S. pneumoniae cells.
SQ Sequence 5816 BP; 1640 A; 1190 C; 1424 G; 1562 T;

alignment_scores:
       Quality: 230.50   Length: 179
       Ratio: 2.077   Gaps: 3
Percent Similarity: 62.011   Percent Identity: 34.078

alignment_block:
US-09-252-691-7056 x v65255

Align seg 1/1 to: v65255 from: 1 to: 5816

45   LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGl 61
     ||||||||||||| ::: |||||::: ||||||:: ||
4945 CTGTTAACAAACACCGGTGTCATTTCCAGTGTAACAGATGACAAGGG 4994

61   yArgSerThrLeuLysAspPheIlePro.....ValGlnGlyValTyrA 76
     ||||:::||||: ||||||:::::||||| :: :||||||
4995 TCGCAAGACGGTTGTCGACCTCTTGCCCAATGTCAAAGAGCGCCATTTACC 5044

76   laAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsn 92
     ::|||||||||| |||:::|||||:||||||||:::||||:||||||
5045 CGTGGGGTCGTTTGGACTGGGATACATCAGGAGTCTTGATTTTGACCAAT 5094

93   AspGlyValLeuGlnAlaAlargLeuThrGlnProGlyLysArgThrGlyLy 109
     ||||| ::: |||||| :::::   ::::: || ||||||
5095 GATGGGGACTTTACAGACAGGATGATTCACCCTCGTAATGAGATTGACAA 5144

109 sIleTyrTyrValGlnInValGluGlyGluProAspAspAlaSerLeuAlaL 126
     |:||||||| :::::|||||||||| : |||||| ::|||
5145 GGTTTATGTCGCGGCGTGTTAAAGGTGTGGCCAATAAGGACAAATCTTCGCC 5194

126 ysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGly 142
     :::::::||||::||:::: ||| ||||||||||||
5195 CCTTGACCCGGTCTTGAGATTGATGGTAAGAAAACCAGCTGTT 5244

143 IleGluArgValAlaAsnGluProGluTrpLeuTrpProArgAsnProProIl 159
     ||| :::: ::||:: |||::::: ||:::::
5245 TATGAAATTCTC...............AAAGTGGACCCAGT 5270

159 eArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrG 176
     :::|||::::||| ::::::::|||:::::::|::
5271 CAAAAATCGC.............TCTGTGGTGCAGTTGACCATCCATG 5305

176 luGlyTyrArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhePro 192
     ||||: ::::||||| :||:::||||: ::: |||
5306 AAGGGCGTAACCATCAGGTAAAAAGATGTTGAGCTGTTGGTCTCCAA 5355

193 ThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSerLe 209
     ::: :::||| ||| ||||:::::||||: ::|||||
5356 GTAGATAAGTTGTCTCGGACTCGTTCGGACACCT.AGACTTGACAGACT 5404

209 uAlaAsnGlyGluTrpArgAspValThrProLysGlu 221
     |||::::: :: :::::: |:
5405 CCGTCCAGGAGAATCCCGTCGTCTTAATAAAAAAGAA 5441

seq_name: N_Geneseq_36:V52369

seq_documentation_block:
ID V52369 standard; DNA; 748 BP.
AC V52369;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:236.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 1257; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 748 BP; 195 A; 191 C; 135 G; 227 T;

alignment_scores:
       Quality: 222.50   Length: 185
       Ratio: 2.079   Gaps: 4
Percent Similarity: 57.838   Percent Identity: 32.973

alignment_block:
US-09-252-691-7056 x v52369/rev

Align seg 1/1 to reverse of: V52369 from: 1 to: 748

45   LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGl 61
     ::::::||||||||||| |||
641 ATGATGAACAAGCCCAAAGGAGTTATCTCAGCGACTGAGGATCCCAAGCA 592

61   yArgSerThr........LeuLysAspPheIleProValGlnGlyValT 75
     |||:::::   :: :|| |||
591 CAGAACCGGTTCTGGACTTGTTGGATGACTTGGCGCCGGGAGTAAGGAAGTTT 542

75   yrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThr 91
     :::||||||||| |||::::::: |||:::|||||||||||
541 TCCCAGTAGGACGCGCTTGGAATTGACACGCATGCGTTTGCTCTTGACC 492

```
  92 AsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGl 108
     ||||||||  |||        |||  |||  :::|||:
 491 AATGATGGTCAGCTGGCTCATGTTCTTTCGCCCAAGCGTCATGTGGA 442

 108 YLysIleTyrTyrValGlnValGluGlyValGluProAspAlaSerLeuA 125
     |||  :  ::|||||||:||||::|
 441 CAAGACTTATCTGGCACAAGTCAAGGGAATCATGACCCAAGAGATGTGG 392

 125 laLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAla 141
     :::  |||||||||||  |:::|||          |||||
 391 AGACATTTGCTGAGGGTATTCCTCCAAGACTTTACCTGTCAACCCGCT 342

 142 GlyIleGluArgValAsnGluProGluTrpLeuTrpProArgAsnProPr 158
     ::|||  |||
 341 ATACTGGAGCTTGTA........................................ 327

 158 oIleArgGluArgLysSerIleProThr...........SerTrpLeuL 171
     |||||||     |||       :::
 326 ............TCCATAGATACAGAAAAGAATCAAAGCCAAATCC 293

 171 ysIleThrTyrGluGlyValGlyArgAsnArgGlnValArgArgMetThrAla 187
     ||||||:::  |||||||:   ::|||:||||||||||||:|||||
 292 GTGTGACCATTGCAGAAGGGAAGTTCATCAGATCAAGCCGTATGGTGGGC 243

 188 HisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTy 204
     :::  |||       ::|||:   |||  ::|||||||
 242 TACTGTGGCAAGGAAGTGGTGACTTGCCAACGATTGACTATGGGAACTTT 193

 204 rThrLeuAsp...SerLeuAlaAsnGlyGluTrpArgAspValThrProL 220
     ::|  |||     :|:|||||||||||||||:||||||:
 192 AGTATTAGATGAGAACCTAGAACGAGGGGAATGGCGTCGCTTGACCAAGG 143

 220 ysGlu 221
     ::|||
 142 AAGAA 138


seq_name: N_Geneseq_36:X13358

seq_documentation_block:
ID   X13358 standard; DNA; 2285 BP.
AC   X13358;
DT   19-MAR-1999 (first entry)
DE   Enterococcus faecalis genome contig SEQ ID NO:421.
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
OS   Enterococcus faecalis.
PN   WO9850555-A2.
PD   12-NOV-1998.
PF   04-MAY-1998; US-066009.
PF   14-NOV-1997; US-044031.
PR   06-MAY-1997; US-044031.
PR   16-MAY-1997; US-046655.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Dillon PJ, Kunsch CA;
DR   WPI; 99-045171/04.
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
PS   Claim 1; Page 1642-1643; 2084pp; English.
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
```

```
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
SQ   Sequence    2285 BP;    672 A;    458 C;    408 G;    743 T;


alignment_scores:
            Quality:  208.00                Length:     183
              Ratio:    1.926                  Gaps:       5
Percent Similarity:   59.016    Percent Identity:  33.880

alignment_block:
US-09-252-691-7056 x X13358/rev

Align seg 1/1  to reverse of: X13358  from: 1 to: 2285

  45 LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGl 61
     :::::|||||||  |||
1855 ATGTTGAATAAGCCAAAAGGTGTGGTTAGCGCTGTCTCAGAT...GCATC 1809

  61 yArgSerThrLeuLysAspPheIleProValGln.............Glyv 74
     ::::|:|||:  |||:  |||
1808 TAAAAAAACAGTTATTGATTTGGCCCACAAGACAGACGCCCCTGGTC 1759

  74 aITyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeu 90
     ::|||    |||||||||||  ||:::||||||||||:::|||
1758 TTTATCCAGTGGCAGACTCGATGCGATACGGAAGGTTACTTCTATTA 1709

  91 ThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgTh 107
     |||  ::|||     :||||     ::|||::|||||
1708 ACAGATAATGGACAATTAGGTATCAATTAATTCGGCCAAATAAAGAAGT 1659

 107 rGlyLysIleTyrTyrValGlnValGluGlyValGluProAspAlaSerL 124
     ::|||  |||        :::
1658 AGCGAAAGTGTATGAAGTGAAAGTCAATGGTTGGTCTCTGCAGAAGATT 1609

 124 euAlaLysLeuArgAsnGlyValThrLeuAsnAspGly...ProThrLeu 139
     :::|||  |||          ||                       |||
1608 GTGCTAAATTTAAAGATGGGATTGTTTTGCAGGAGCGCATTCAATGTAAG 1559

 140 ProAlaGlyIleGluArgValAlaAsnGluProGluTrpProLeuTrpProArgAs 156
     ||||||  ::|||
1558 CCAGCAAAAATAACTGTTCTTGCAGCGACTGAA.................... 1526

 156 nProProIleArgGluArgLysSerIleProThrSerTrpLeuLysIleT 173
     |||  :::  :||||
1525 .....................ACGGAGAGTCATGTTTTCCTGA 1504

 173 hrLeuTyrGluGlyValArgAsnArgGlnValArgArgMetThrAlaHisVal 189
     ||:::   |||||||||||||||:|||:||||||||||
1503 CTATTCAAGAGAGGAAAATTCACCAAGTGAAAAAAATGTTTTTATCCGTG 1454

 190 GlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLe 206
     |||        |||   :::|||||:::|||::||||
1453 GGGAAAAAGTGACTGCTTTAAAACGCCAAACTATGGGGCCGTTGCGGCT 1404

 206 uAsp...SerLeuAlaAsnGlyGluTrpArgAspValThrProLysGlu 221
     |||        :::|||    ||           :::|||    ::|||
1403 AGATCCGCAATTACCTTTAGGTGCTTATCGTTCATTAACAAGAGAAGAG 1355


seq_name: N_Geneseq_36:V22799

seq_documentation_block:
ID   V22799 standard; DNA; 3061 BP.
AC   V22799;
DT   24-JUL-1998 (first entry)
DE   Nucleotide sequence of the E3 fragment encoding the PhaG protein.
KW   E3 fragment; Pseudomonas putida KT2440; polyhydroxyalkanoic acid; PHA;
KW   pha G gene; fatty acid biosynthesis; PHA production; rhlA gene product;
KW   P aeruginosa PG201; qin gene; breakage; ester bond; complementation; ss.
OS   Pseudomonas putida.
FH   Key             Location/Qualifiers
FT   CDS             911..1798
FT                   /*tag= a
```

PN WO9806854-A1.
PD 19-FEB-1998.
PF 13-AUG-1997; U14388.
PR 13-AUG-1996; US-700576.
PA (MONS) MONSANTO CO.
PI Kruger N, Steinbuchel A;
WPI; 98:-159545/14.
DR P-PSDB; W56310.
PT PhaG gene involved in poly:hydroxy:alkanoate production - identified
PT by phenotypic complementation in Pseudomonas putida mutants
PS Claim 1; Pages 42-43; 57pp; English.
CC The present sequence represents the E3 fragment of Pseudomonas putida
CC KT2440 pha (polyhydroxyalkanoic acid) G gene locus. The sequence contains
CC three open reading frames (ORFs). ORF1 and ORF3 are localised only
CC incompletely on this fragment, with ORF1 lacking the 5'-region, and
CC ORF3 lacking the 3'-region. ORF2 encodes PhaG. The PhaG protein is
CC involved in the linkage between fatty acid biosynthesis and PHA
CC production when P. putida is grown on a simple carbohydrate source
CC such as gluconate. The PhaG protein shows a 44% overall identity to the
CC rhlA gene product, of P. aeruginosa PG201. The C-terminal regions of
CC RhlA and PhaG show high homology to a qin gene region of P. aeruginosa,
CC which encodes a quinolone sensitivity protein. Pha also has a region
CC which shows homology to a group of enzymes, most of which are hydrolases
CC active on esters such as lipids. PhaG may therefore be involved in thr
CC breaking of ester bonds. The phaG DNA is used to genetically transform
CC plants e.g. maize, potato, sugar beet, tobacco, wheat or high oil seed
CC plants chosen from soybean, canola, rape seed, sunflower, flax and peanut
CC and Arabidopsis, which can then be harvested. Synthesis of PHA from
CC gluconate in mutant microorganisms can be restored by complementation
CC using the phaG DNA.
SQ Sequence 3061 BP;   615 A;   980 C;   863 G;   603 T;

alignment_scores:
   Quality: 206.00   Length: 189
   Ratio: 1.856   Gaps: 6
Percent Similarity: 58.730   Percent Identity: 32.275

alignment_block:
US-09-252-691-7056 x V22799   .
Align seg 1/1 to: V22799 from: 1 to: 3061

 39 GlnProThrArgValIleLeuPheAsnLysProTyrAspValLeuProGl 55
    :::::::::::::::: ::::::::   :::   :::
 46 CGCCCGGCCCGCTACCTGTACAAGCCCACTGGCTGCGTAACGGC   95

 55 nPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIleProVal. 71
    : ::::::::   :::::::: :::   :::::
 96 CACCCACGATCCGCAACACCGTACCGTTCTC...GACCTGCTGCCAGCGG 142

 72 ....GlnGlyValTyrAlaAlaGlyValAspArgLeuAspArgAspArgSerGlu 85
       ::::: ::::::::::::::::::::::: :::::::
143 CGTTGGCGAGATGACCTGCACATAGCCGGGCGCCTGGACTTCAACACCACC 192

 86 GlyLeuLeuValLeuThrAsnAspAspGlyValLeuGlnAlaAlaArgLeuThrGl 102
    :::::::::::::::::::::::::::::   ::   :::
193 GGCCTGATGATCCTGACCAACGATGGCCAATGGTCACGGCGGCCTGACCA 242

102 nProGlyLysArgThrGlyLysIleTyrrTyrValGlnValGlyValGlyLeuuP 119
    :::: :::   ::::::::::: :::::   :::
243 GCCTGCCACCAAGCTGCCCAAGCATTATCTGGTGGACACCGAGGACCGAGA 292

119 roAspAspAlaSerLeuAlaLysLeuuArgAsnGly......ValThrLeu 133
    :::::::::::::::::::: ::::::::::::   :::::::
293 TTGGCGAGCACTATGTGGCCAAGTTTCCGCGAGGGTTTCTATTTTGCCTTC 342

134 AsnAspGlyProAlaGlyIleGluGluArgValAsnGluProGl 150
    :::::: ::: ::::::::::::::::   :::::::
343 GAAGACCTCACCACCCACCTGCCCAGCTGGACATCCTCGGC........ 384

150 uTrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProT 167
    ::: :::   :::

385 ....CCCCACCGAGCCCGG............ 399

167 hrSerTrpLeuLysIleThrLeuTyrGluGlyAsnArgGlnValArg 183
    ::::::::   :::::::::::::::::: ::::::::::
400 ....................CTGGCGATCGTCGAGGGCGTTACCACCAGGTCAAG 435

184 ArgMetThrAlaHisValGlyPheProThrLeuArgLeuuIleeArgTyrAl 200
    ::::: ::::::: :::::::::::   :::   :::
436 CGCATGGTTCGGCCATTTCAACACACAAGGTGATCGGGCTGCATCGGGAGAG 485

200 aMetGlySerTyrThrLeuAsp...SerLeuAlaAsnGlyGluTrpArgA 216
    :::::::::::   ::::::::::: ::::::::::::   :::::::::_
486 CATGGGGGCGATCCGGCTGGATCCGGGGTTGGCGCCGGGAGTATCGTG 535

216 spValThrProLysGlu 221
    :::::::::: :::::::
536 AACTGACGCCAATGAG 552

seq_name: N_Geneseq_36:V74482

seq_documentation_block:
ID V74482 standard; DNA; 1126 BP.
AC V74482;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #171.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key   Location/Qualifiers
FT misc_feature 1141..1200
FT   /*tag= a
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT misc_feature 2941..3000
FT   /*tag= b
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT misc_feature 4741..4800
FT   /*tag= c
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT misc_feature 6541..6600
FT   /*tag= d
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT misc_feature 8341..8400
FT   /*tag= e
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
FT misc_feature 10141..10200
FT   /*tag= f
FT   /note= "these bases represent a line of missing text in
FT   the sequence listing in the specification. They
FT   are included to maintain the nucleotide numbering
FT   given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 833-839; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence   11126 BP;   3428 A;   1959 C;   1551 G;   3822 T;

alignment_scores:
    Quality: 202.50   Length: 219
    Ratio: 1.633   Gaps: 7
Percent Similarity: 56.621   Percent Identity: 29.224

alignment_block:
US-09-252-691-7056 x V74482/rev

Align seg 1/1 to reverse of: V74482 from: 1 to: 11126

     7 IleThrProGluAsnThrMetThrLysThrSerPheArgLysHisArgVa 23
       ||| ||||||||| ||| :| ||| :: ::
  6104 ATAGAACCAGAAAAC......GATAAGATAACAGTTCGTGGAGAATTAAT 6061

    23 lGluArgPheSerSerArgGlnAlaThrArgArgThrProGluProGlnP 40
       :||| ::
  6060 AGAATACATTGAAAT.......................... 6045

    40 roThrArgValIleLeuPheAsnLysProTyrAspValLeuProGlnPhe 56
       |||::||:::|||||||| |: :: |||
  6044 ....GTCTATATATGTTAAATAGCCAAAGGTTATATA...TCAGCA 6003

    57 ThrAspGluAlaGlyArgSerThrLeuLysAspPheIleProValGln.. 72
       |||:::::   ::|||:|| |||||||||||||||:|||||
  6002 ACTGAAGATCATCATTCAAAAACCGTTAATGATTTAATACCTGAATACCA 5953

    73 ....GlyValTyrAlaAlaGlyLeuArgLeuAspArgAspSerGluGlyLyL 87
       ::|||||||:|||||||||||||||||||
  5952 ACATCTAAATATTTTTCCGGTTGGTCGTCTTGATAAAGATACAGAAGGGC 5903

    87 euLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnPro 103
       ||||:::::::||||||||| :||:|| :: :|::
  5902 TTTTATTGATAACGAATGATGGTGATTTAATCATGAATTAATGAGTCCA 5853

   104 GlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGluProAs 120
       ::||||:||||||||:||| ||| :|
  5852 AACAAACATGTTCTAAAAAATATGAAGTTATTTCAGCAAATCCTATCAC 5803

   120 pAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyP 137
       |||||||||||||||||||||||||||||:|||||:|:||
  5802 TGAAGACGACATTCAAGCTTTTAAAGAGGGTGTAACACTGACTGATGGCA 5753

   137 roThrProAlaGlyIleGluArgValAlaAsnGluProGluTrpLeuTrp 153
       ::| :::   :::
  5752 AGGTTAAGCCCTGCAATATTAACTTAATATTGATAAT............. 5718

alignment_block:
US-09-252-691-7056 x X13710/rev

Align seg 1/1 to reverse of: X13710 from: 1 to: 1081

    45 LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGl 61
       :::::::||||||

PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1505-1506; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence 676 BP; 163 A; 136 C; 77 G; 236 T;

alignment_scores:
    Quality: 201.50   Length: 180
    Ratio: 2.056   Gaps: 2
Percent Similarity: 54.444   Percent Identity: 31.111

alignment_block:
US-09-252-691-7056 x V74878/rev

Align seg 1/1   to reverse of: V74878   from: 1   to: 676

 44 IleLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAl   60
    : ::::: : ::: ::: :: :::: :::: :: : :: :::::::
537 ATTTTMTTCCATAAMCCAAMCTCAAGTGATMACAGTGTATCTGATGATGAG 488

 60 aGlyArg......SerThrLeuLysAspPheIleProValGlnGlyValT   75
    |||||| :::::
487 AGGACGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 438

 75 yrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThr   91
    ::::: ||| ||:::::::::::::||||||:::||||||||||
437 NNNNNNNNNNNNNGACTATGAATACTTCTCGGATTGTTATTACTCACA   388

 92 AsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGl 108
    ::::: ||||| || ||||||::::|||||:::||||:|:::||||
387 AATGATGGTGAATTTACTAATTAATGACACATCCAAGATATCAAATTAA 338

108 yLysIleTyrTyrValGlnValGluGluProAspAspAlaSerLeuA   125
    ::::: |||| :::::|||:||||::||||||||||||||
337 GAAAAATATGTTGCAAATTAAAAGGTTATTTAATGAGAGAAGAGTGA   288

125 laLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAla 141
    ||:::|||||||||| |||:::||||| ||| |||||||
287 AAGCGCTAGAAAAAGTATTGAATTAGAAGATGGTATGACGCAACCGGCT   238

142 GlyIleGluArgValAsnGluProGluTrpLeuTrpProArgAsnProPr 158
    ::::: :::::||||::||:||||||::::||::
237 GAAGTGAAA..GT 227

158 oIleArgGluArgLysSerIleProThrSerTrpLeuTyrLeuSerTleLeuT 175
    :::::::|||:::::::::::::||||||||||||||||
226 TAAGAAACAAGATAAAGATAAAAACACCAACATTGGTTGAAATTACAATTA 177

175 yrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhe 191
    |||||||||||||||||||||||||||||||||||
176 CAGAAGGTCGTAATCGCCAAGTCAGAAGAAGAATGTTCGAACATTTGGTCAT 127

192 ProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrLeuAspSe 208
    ::: :::||| |||
    :::::: ::   :::

--- (left column) ---

536 ATGTTGCATAAACCACAAGGTGTCATT...TCTGCGACGAAGATCAGCG 490
 61 yArgSerThrLeuLysAspPheIleProValGlnGlyValValTyrAla.... 76
    ||||| |||
489 AGATGAAACGGTAATTGACTTGTTAACTGATCAGGATTMCCGTGCCGACT 440
 77 ....AlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeu 90
    ::::::||||||||||||||||||:::
439 TGTCCCAGTTGGTCGCTTGGATWAAGRTMCCGARGGCTTGCTTCTACCA 390
 91 ThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgTh 107
    |||||||| ||| :::||| :::::|| |||:::::
389 ACAAATGATGGCCAGTTAGCCGCATCAATTATTATCMCCCAAAAGCATGT 340
107 rGlyLysIleTyrTyrValGlnValGluGly......GluProAspAspA 122
    :::::||||||||||:::::|||||||| ::|||
339 GGATAARGAWTATTTAGCTGAAGGACTTGTGCACACGGAAGATT 290
122 laSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
    :::::|| ::: :::||:::||||||||||
289 GTCAACTG.....TTTGCCAAGGATTTGCCTTGAAAGATGGAGAAATG 246
139 LeuProAlaGly.......IleGluArgValAsnGluProGluTrppLe 152
    :: ::: |||||| ::|||||| |||
245 GTTAAACCAAGTCGTTAGTCAGTCATTGAGACGATCAATGAGGAAGAA.... 201
152 uTrpProArgAsnProProIleArgGluArgLysSerIleProThrSerT 169
    :::::|||||||||||| |||||||
200 ....GGGACTTCCG 191
169 rPLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMet 185
    ::::::|||||||||| |||:::||||||||||||
190 TAGTCCGTTTAGTGATTCAAGAGGGGAAATTTCATCAAGTCAAACGGATG 141
186 ThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGl 202
    ||| |||||| |||::: ||| ||| |||||
140 TTAGGCCACTGTAGGCAAACCTGTGACGTATTTAAAAACGATTACGAATGGG   91
202 ySerTyrThrLeuAsp...SerLeuAlaAsnGlyGluTrpArgAspValT 218
    ::::::||||| ||| ::: ||||||| :::
 90 CGAACTATGGTTGGATGAAAACGTTAGAACTAGGGGAATATCGGCCCCTTAA   41
218 hrProLysGlu 221
    || |||
 40 CAGAAGCCGAA 30

seq_name: N_Geneseq_36:V74878

```
126 CAAGTGACGAAACTGTCTCGTATTGAATATGGGCCTTTAAATGTTGTCGG 77

208 rLeuAlaAsnGlyGluTrpArgAspValThrProLysGlu 221
    :|||  ||||||  |||  ::||||||| |||
76  TTTGAATGCTGGTGAAGGACGTGTATTGACGGCCTCATGAA 37
```

seq_name: N_Geneseq_36:V52353

```
seq_documentation_block:
ID   V52353 standard; DNA; 4692 BP.
AC   V52353;
DT   23-OCT-1998 (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:220.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI; 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 1230-1232; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, or expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence 4692 BP; 1429 A; 1008 C; 820 G; 1434 T;
```

```
alignment_scores:
    Quality:       185.50         Length:    189
    Ratio:         1.718          Gaps:      7
Percent Similarity: 57.143        Percent Identity: 30.688
```

```
alignment_block:
US-09-252-691-7056 x V52353

Align seg 1/1 to: V52353 from: 1 to: 4692

39   GlnProThrArgValIleLeuPheAsnLysProTyrAspValLeuProGl 55
     ::||:|| :|:::|| |||||| :::   ||||||
3072 AAACCTGCTGGTGCCGTTACAGCCAACAAAGACCAAGGAA..CTTCCG.. 3116

55   nPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIlePro.... 70
     ||||||:|:|:||
3117 ..........ACCCTCATCGACCTGCTTCCCATCTA 3141

71   .ValGlyGly.....ValTyrAlaAlaGlyLysArgLeuAspArgAspSer 84
```

```
3142 ACATCCAGTCGACAGCTCTATGCCGTTGGCCGACTGGACCGAGATACA 3191
     ::||||| ::|||||||||||||||||:::||||||||||||||||::::
85   GluGlyLeuLeuValLeuThrArgAsnAspGlyValLeuGlnAlaArgLeuTh 101
     |||||||:|||||  :||||||||||| |:|||||
3192 ACGGGACTCCTCTCTTGACCGATAACGGTCCCTTGGGCTTCAGCTCCT 3241

101  rGlnProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGlyl 118
     ::||||  |||  ||||:|::||||:::||::|||
3242 CCATCCCAATATCATGTCGATAAGACTACCAAGTTGAGGTAATTGAGGGAC 3291

118  luProAspAspSerLeuAlaLeuArgAsnGlyValThr...Leu 133
     :::  :::||||||||::|||:|||
3292 TTCTAACACCTGACCATATCCAAACCTTTCAAAAAGGAATTGTCTTTTA 3341

134  AsnAspGlyProThrGluProAlaProAlaGlyIleGluArgValAsnGluProGl 150
     ::|||  ||||  :||||  ||::|
3342 GATGACACTGTCTGTAAACCCGCAAAACTAGAGATTCTATCT......... 3383

150  uTrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProT 167
     |||
3384 ..........................GCAAGTCSCTCCC 3396

167  hrSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArg 183
     |||  :::||||||| |||||||||::::::|:|
3397 TCAGTCAAGCCCTATCACCATTCAGAAGGAAAATTTCATCAAATCAAG 3446

184  ArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAl 200
     ::|||  ||||:||  |||::||||||||||||||
3447 AAAATGTTCCTCTCGGTTGGTGTTAAGGTGACTAGCCTCAAAAGAATCCA 3496

200  aMetGlySerTyrThrLeuAsp...SerLeuAlaAsnGlyGluTrpArgA 216
     :::||:||||||| :::|||||||:|:|||
3497 ATTTGGGGACTTCACATTGAACCCAGATTTAGCAGAAGTAACTACCGCC 3546

216  spValThrProLysGlu 221
     ||||||
3547 CTTTGAACCAAAAGAG 3563
```

seq_name: N_Geneseq_36:V52508

```
seq_documentation_block:
ID   V52508 standard; DNA; 720 BP.
AC   V52508;
DT   23-OCT-1998 (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:375.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997; U19588.
PR   31-OCT-1996; US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI; 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 1388; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded.
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
```

CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, or expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence 720 BP;   219 A;   136 C;   176 G;   189 T;

alignment_scores:
     Quality: 167.00   Length: 98
     Ratio: 2.530   Gaps: 1
Percent Similarity: 67.347   Percent Identity: 35.714

alignment_block:
US-09-252-691-7056 x V52508

Align seg 1/1 to: V52508 from: 1 to: 720

 45 LeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGl 61
     ||||:||||||||| ::||||||| :: ::||||:|| |:||
332 CTGCTTAACAAACCACGCGGTGCGTGATTCCAGTGCAGATGATAAGGG 381

 61 yArgSerThrLeuLysAspPheIlePro......ValGlnGlyValTyrA 76
     |||:::|||:||| |||| :: ::|||
382 TCGCAAGACGGTTGTCGACCTCTTGCCCAATGTCAAAGAGCGTATTTACC 431

 76 laAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsn 92
     ::|| ||| ::||||||||||:::||||||||
432 CTGTGGGTCGTTGGACTGGGATACATCAGGTGTCTTGATTTGACCAAT 481

 93 AspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLy 109
     ||||| ::||: ||||| ::||||
482 GATGGGGACTTTACAGACGAGATGATTCACCCTCGTAATGAGATTGACAA 531

109 sIleTyrTyrValGlnValGluGluProAspAspAlaSerLeuAlaL 126
     |:::|| |||| :||||||||:|| ::|||
532 GGTTTATGTCCGCGGTGTTAAAGGTGGCCAATAAGGACAATCTCCGCC 581

126 ysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuPro 140
     ||| ::|||| ||| |||
582 CCTTGACCCGGTGGTCTTGAGATTGATGGTAAGAAAAACCAAGCCA 625

seq_name: N_Geneseq_36:X20248_07

seq_documentation_block:
Continuation (8 of 10) of X20248 from base 700001 (Borrelia burgdorferi polynunucleotide s
WP Sequence split into 10 fragments   LOCUS X20248 Accession X20248
WP   Fragment Name   Begin   End
WP   X20248_00   1   110000
WP   X20248_01   100001   210000
WP   X20248_02   200001   310000
WP   X20248_03   300001   410000
WP   X20248_04   400001   510000
WP   X20248_05   500001   610000
WP   X20248_06   600001   710000
WP   X20248_07   700001   810000
WP   X20248_08   800001   910000
WP   X20248_09   900001   910715

alignment_scores:
     Quality: 151.00   Length: 224
     Ratio: 1.238   Gaps: 6
Percent Similarity: 54.464   Percent Identity: 22.768

alignment_block:
US-09-252-691-7056 x X20248_07 ..

Align seg 1/1 to: X20248_07 from: 1 to: 110000

  2 IleMetArgGlnIleLeuThrProGluAsnThrMetThrLysThrSerPh 18
     ||| :|||||||:|| :|||||||:||||:||| :::
85250 ATAAGAAGAAACTTGAAGAGTTAATACACTATTGCTAAGCTTGGGGA 85299

 18 eArg...........LysHisArgValGluArgP 26
     ::: ||| ::::::||||
85300 TAAAGTAACTTAGGAGACAGAATAATTATAAAAACAGATTTTGTTT 85349

 26 heSerSerArgGlnAlaThrArgArgThrProGlnProGlnProThrArg 42
     || ||:::||| ||:
85350 TTAAAGATTTCAAATTATATAATAGAATT............... 85378

 43 ValIleLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGl 59
     :::||||||| :: ::|||||||||| ::::|||
85379 TATCTAGCTCTTAACAAGCCTAGAAATTATTATGTTCAATTTGATGT 85428

 59 uAlaGlyArgSerThrLeuLysAspPheIle....ProVal......GlnG 73
     ||||||| :::::::|:|| || :: :||
85429 TGATGGAGAAAGTTAGCAATATCTTTGGTTCAGCCCTTATTTAAAGAGC 85478

 73 lyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuVal 89
     |||:::||||||||| :|| :|||||||||:||||||||
85479 GTGTGTTTCAATTGGTAGCCTTGATTTTAAAAGCTCTGGACTTTTATTA 85528

 90 LeuThrAsnAspGlyValLeuGlnAlaAlaArgLeuThrGlnProGlyLysAr 106
     :::|||||||| :||:|||||| |||||||||
85529 TTCACCAATGATGGTAAATTTGCAAACGATATTATTCATCCAAGGCAAAA 85578

106 gThrGlyLysIleTyrTyrValGlnValGluGlyLeuProAspAspAlaS 123
     ::::: ||| ::: :|| || |||
85579 AGTTGAAAGAGAATAATATTATTGAATCAAAAAAAGATATTGATGAAAATT 85628

123 erLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeu 139
     ||| |||||||||:::||||||:||:|||
85629 TGCTTATTTCTTTTAAATCGGGTATAAAGGTA.............. 85660

140 ProAlaGlyIleGluArgValAsnGluProGluTrpProLeuProArgAs 156
     ::::::||||||||
85661 ..............AAAAAGAGATTTTTTAAATTAAAATC 85686

156 nProProIleArgGluArgLysSerIleProThrSerTrpLeuLysIleT 173
     ||| ::||||:|||||||||:|
85687 TTATGAAATTTAAATAAAAATTCT..............GCTAGAATGA 85721

173 hrLeuTyrGluGlyLeuArgAsnArgGlnValArgArgMetThrAlaHisVal 189
     ||| ||||:||| ::||||||||||||
85722 TTTTAGATGAAGGGAAAATAAGCAAAAAAGTGTTTTTGAGCAAG 85771

190 GlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrIle 206
     ::: ||||||||||||||:::|||
85772 AATATTTTTTAAAAAAATTCATAGAATTAGAATTGGCAAATATTAATTT 85821

206 uAspSerLeuAlaAsnGlyGlu 213
     ||||||||| :::||||||
85822 AGATAGTTTAAAGGAGGGTCAA 85843

seq_name: N_Geneseq_36:T84021

seq_documentation_block:
ID   T84021 standard; DNA; 584 BP.
AC   T84021;
DT   01-SEP-1998 (first entry)
DE   DNA encoding a Staphylococcus aureus protein of unknown function.
KW   Staphylococcus aureus protein; antisense sequence; control;
KW   Staphylococcal gene; regulatory element; bacterial gene expression;
KW   vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW   toxic shock syndrome; ss.
OS   Staphylococcus aureus.
FH   Key   Location/Qualifiers
FT   CDS   complement (65..511)

```
FT   WO9730070-A1.   /*tag- a
PN   WO9730070-A1.
PD   21-AUG-1997.
PF   19-FEB-1997.
PR   20-FEB-1996; US-011888.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR   WPI: 97-424969/39.
DR   P-PSDB; W28082.
PT   Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
     to isolate antimicrobial compounds, and in vaccines against S.
     aureus infection
CC   Claim 9; Page 823; 989pp; English.
CC   The present sequence encodes a Staphylococcus aureus protein of
CC   unknown function. The present sequence was isolated from a
CC   library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
CC   sequence can be used in the construction of ribozymes and antisense
CC   sequences to control the expression of Staphylococcal genes. The DNA
CC   sequence is also useful as a source of regulatory elements for the
CC   control of bacterial gene expression. The encoded protein may be used
CC   to produce vaccines to enable a host to produce specific antibodies
CC   with antibacterial action. These vaccines and antibodies would protect
CC   a host against invasion by S. aureus, and conditions relating to
CC   Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC   skin syndrome, and toxic shock syndrome.
SQ   Sequence 584 BP; 167 A; 109 C; 84 G; 221 T;

alignment_scores:
        Quality:  128.50          Length:  123
        Ratio:    1.736           Gaps:    6
Percent Similarity: 60.163     Percent Identity: 32.520

alignment_block:
US-09-252-691-7056 x T84021/rev  ..

Align seg 1/1  to reverse of: T84021  from: 1 to: 584

7   IleThrProGluAsnThrMetThrLysThrSerPheArgLysHisArgVa  23
    |||  ||||||||||        |||  |||   ||
391 ATAGAACCAGAAAAC......GATAAGATAAACAGTTCGTGGAGAATTAAT  348

23  1GluArgPheSerArgArgGlnAlaThrArgArgThrProGluProGlnP  40
    :|||   :::::::::
347 AGAATACATTGAAAAT..................................  332

40  roThrArgValIleLeuPheAsnLysProTyrArgAspValLeuProGlnPhe  56
       ||:|||||||||||            |||  ::
331 ....GTCTATATTATGTTAAATAAGCCAAAAGGTTATATA...TCAGCA  290

57  ThrAspGluAlaGlyArgSerThrLeuLysAspPheProValGln..  72
    ::|||:::|||     ||:||:||||||
289 ACTGAAGATCATCATTCAAAAACCGTTATTGATTTAATACCTGAATACCA  240

73  .....GlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLyL  87
            |||||||||:|||  ||  ||||||::||||
239 ACATCTAAATATATTTTCCGGTTGGTCGTCTTGATAAAGATACAGAAGGGC  190

87  euLeuValLeuThrAsnAspGlyValValLeuGlnAlaArgLeuThrGlnPro  103
    |||||||||||||||   |||  ::|||:::  :|||||
189 TTTTATTGATAACGAATGATGGTGATTTAATCATGAATTAATGAGTCCA  140

104 GlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGluPro..  119
    ::|||:|||||||  ||| |||  |||:::|||
139 AACAAACATGTTTCTAAAAAAATAGAAGTT...ATTTCAGGAAATCCTAT  93

120 ......AspAspAlaSer  123
          ||||||||||
92  CACTGAAGACGACAATTCA  74

seq_name: N_Geneseq_36:V65214
```

```
seq_documentation_block:
ID   V65214 standard; DNA; 3071 BP.
AC   V65214;
DT   24-DEC-1998  (first entry)
DE   DNA encoding a S. pneumoniae protein of unknown function.
KW   Streptococcus pneumoniae protein; recombinant; gene expression;
KW   DNA chip; virulence; antibody; infection; detection; treatment; ss.
OS   Streptococcus pneumoniae.
PN   WO982607-A1.
PD   18-JUN-1998.
PF   09-DEC-1997; U22578.
PR   13-DEC-1996; US-036281.
PA   (ELIL ) LILLY & CO ELI.
PI   Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI   Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR,
PI   Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI   Young Bellido ML;
DR   WPI: 98-348529/30.
DR   P-PSDB; W80138.
PT   Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
     for evaluating gene expression, and identification of virulence
     genes
PS   Claim 1; Pages 52-54; 333pp; English.
CC   This sequence encodes a Streptococcus pneumoniae protein of unknown
CC   function. The invention provides DNA sequences (v65201 to v65304) from
CC   the Streptococcus pneumoniae genome and corresponding protein sequences
CC   (W80605 to W80728). A recombinant host containing a vector comprising any
CC   of the above nucleic acids can be used for the recombinant expression of
CC   the protein sequences. The invention also provides a DNA chip having
CC   arrayed on it at least 15 base pair fragment of any one or more of these
CC   DNA sequences. The DNA chip can be used methods for evaluating gene
CC   expression in S. pneumoniae and for identifying virulence genes in
CC   S. pneumoniae. Antibodies that selectively bind to the above proteins or
CC   peptide fragments can be used to treat S. pneumoniae infection. The
CC   antibodies can also be used to detect S. pneumoniae cells.
SQ   Sequence 3071 BP; 886 A; 588 C; 668 G; 929 T;

alignment_scores:
        Quality:  127.00          Length:  278
        Ratio:    0.984           Gaps:    13
Percent Similarity: 46.403     Percent Identity: 24.101

alignment_block:
US-09-252-691-7056 x V65214  ..

Align seg 1/1  to: V65214  from: 1  to: 3071

5   GlnLeuIleThrProGluAsnThrMetThrLysThrSerPheArgLysHi  21
    :::||::::||||||  |||     :::||||||  :::|||  |||||||
805 GAATTATCGCAGATGAACATGTCAAGGTTAAGACCTTTTAAAAAAGCA  854

21  sArgValGluArg.................................  25
    :::||||||
855 CGAGGTTTCTAAGGGATTGCTGGCCAAGATTAAGTTTCGAGGTGAGCTA  904

25  ..........................................  25
905 TTCTGGTCAATAATCAACCGCAAAAATGCAACGTATCATTGGACGTTGGA  954

26  ..........................PheSerSerArgG1  30
    :::|||  :::
955 GACTACGTTACCATTGACATTCCGCTGAGAAAGGCTTTGAAACCTTGGA  1004

30  nAlaThrArgArgThrProGlu.....ProGlnProThrArgValIleL  45
    |||                          |||     :::|||
1005 GGCTATTGAGCTTCCATTAGATATTCTCATCAGGATGACCACTTCTAG  1054

45  euPheAsnLysProTyrArgAspVal.....LeuProGlnPheThrAspGlu  59
    ::|||::  ||                     :::||  :|:
1055 TCTTGAATAAACCCTATGGAGTGGCTTCTATTCCTAGTGTTAATCACTCT  1104

60  AlaGlyArgSerThrLeuLysAspPheIle.................  69
```

```
CC   genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC   by a process comprising: (a) screening a genomic DNA library using as a
CC   probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC   to 391, identifying members of the library which contain sequences
CC   that hybridise to the target sequence and isolating the nucleic acid
CC   molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC   from an organism, amplifying nucleic acid molecules whose nucleotide
CC   sequence is homologous to amplification primers derived from the
CC   fragment of the S. pneumoniae genome to prime the amplification and
CC   isolating the amplified sequences. The computer readable medium can be
CC   used in a computer-based system for identifying fragments of the
CC   S. pneumoniae genome of commercial importance, or expression modulating
CC   fragments of the S. pneumoniae genome. Products from the present
CC   invention can be used in diagnosis kits and assays, and pharmaceutical
CC   compositions and vaccines for S. pneumoniae.
SQ   Sequence 20199 BP;   3655 C;   6009 A;   4458 G;   6077 T;

alignment_scores:
         Quality:   126.00        Length:      278
         Ratio:       0.992       Gaps:         13
Percent Similarity:  45.683   Percent Identity:  24.460

alignment_block:
US-09-252-691-7056 x V52139

Align seg 1/1  to: V52139  from: 1  to: 20199

    5  GlnLeuIleThrProGluAsnThrMetThrLysThrSerPheArgLysHi   21
       :::::|::|||||::  ||||:|::::::: ::||||||    ::|||||
17566  GAATTTATCCGAGATGAACATGTCAAGGTTAAGAACCTTCTTAAAAAAGCA   17615

   21  sArgValGluArgPheSerSerArgGlnAlaThrArgArg...........   34
       |||||||   |||||||||
17616  CGAGGTTCTAAGGGATTGCTGGCCAAGATTAAGTTCGAGGTGGAGCTA     17665

   35  ...........ThrProGlnProThrArgValIleLeu.............   45
                :::  ::|||||:||  ||||||||::
17666  TTCTGGTCAATAATCAACCCGCAAAATGCAACGTATCTATTGGACGGTTGGA  17715

   45  ..................................................   45
17716  GACTACGTTACCATTGACATTCCCGCTGAGAAAGGCTTTGAAACCTTGGA    17765

   45  ..................................................   45
17766  GGCTATTGAGCTTCCATTAGATATTCTCTATGAGGATGACCACTTTCTAG   17815

   46  ..PheAsnLysProTyrAspVal......LeuProGlnPheThrAspGlu   59
       ::||||||||||||||  |||          :||:::    :||||||||
17816  TCTTGAATAAACCCATGGAGTGGCTTCTATTCCCAGTGTCAATCACTCT    17865

   60  AlaGlyArgSerThrLeuLysAspPheIle.................    69
       :::||||||
17866  ..........AATACCATTGCCAATTTTATCAAGGGTTACTATGTCAAGCA   17906

   70  .....ProValGlnGlyValTyrAlaAlaGlyLeuAspArgAspS       84
       :::
17907  AAATTATGAAAATCAGCAGGTTCACATTGTTACCAGACTAGTAGGGATA   17956

   84  erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu  100
       ::::|||||||  ||||||||| ||||||:|||||||||:||||:||:||
17957  CTTCTGGCTTGATGCTCTTGCCAAGCACGGTTATGCCCATGCACGATTA    18006

  101  ThrGln.....ProGlyLysArgThrGlyLysIleTyrTyrValGlnIva  115
       :::                |||  |||::|::  |||  |||::::::: ||
18007  GACAAGCAGTTGCAGAAGAAATCATCGAGAAACGCTACTTTGCTTTGGT   18056

  115  lGluGly.........GluProAspAlaSerLeuAla.LysLeu        127
       |::||||          |||  ||||:|||||||:::  ||:||
18057  TAAGGGAGATGGACATTTGGACGCAGAGGGAAATTATTGCTCCGATTG    18106

  128  ArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGl  144
```

```
1105  ..........AATACCATTGCCAATTTTATCAAGGGTTACTATGTCAAGCA   1145
            :::||||::  ::||||
  70  .....ProValGlnGlyValTyrAlaAlaGlyLeuAspArgAspS        84
        :::|||   ||| :::  :::   ||||||||||||
1146  AAATTATGAAAATCAGCAGGTTCACATTGTTACCAGACTAGTAGGGACA    1195
  84  erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu  100
      ::::|||||||||||  :::::::::   |||  :::|||||:||||||
1196  CTTCTGGCTTGATGCTCTTGCCAAGCACGGTTATGCCCATGCACGATTA   1245
 101  ThrGln.....ProGlyLysArgThrGlyLysIleTyrTyrValGlnIva  115
      :::|||||       |||  |||   ||
1246  GACAAGCAGTTGCAGAAGAAATCATCGAGAAACGCTACTTTGCTTTGGT   1295
 115  lGluGly.........GluProAspAlaSerLeuAla.LysLeu        127
      |:::|||        :::
1296  TAAGGGAGATGGACATTGGAGCCAGAGGGAAATTATTGCTCCGATTG     1345
 128  ArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGl  144
      |||       :::             |||  ||||||
1346  CGC........GTCATGAAGATTCCATTATTACCA.........       1372
 144  uArgValAsnGluProGluTrpProArgAsnProProIleArgG        161
      |||||||            :::|||
1373  ..........GACGAGTGGCTAAGGGCGGAAAGTATGCCCATACTTC     1409
 161  luArgLysSerIleProeThrSerTrp.............LeuLysIle   172
      ::  :::||
1410  ATAC...AAGATTGTAGGTTCTTATGGAAATAATTCACTTGGTCTATATT   1455
 173  ThrLeuTyrGluGlyValArgAsnArgGlnValAlaArgMetThrAlaHisVa  189
      ||| |||           :::  :::|||
1456  CACCTGCCACACTGGTCGAACCCATCAAAATCCGAGTCCCATTTTCTCATAT  1505
 189  lGlyPheProThrArgLeuIleArgTyrAlaMetGlySerTyrThrL    206
      |||||||||
1506  CGGTTTCCTTGCTGGGA.........                           1524
 206  euAspSerLeuAlaAsnGlyGluTrpArgAsp                216
      :::||:::||    ||| :::|||
1525  ..GATGATTTGTATGGTGGTAGTCTGGAAGAT  1554
```

```
seq_name: N_Geneseq_36:V52139

seq_documentation_block:
ID   V52139 standard; DNA; 20199 BP.
AC   V52139;
DT   23-OCT-1998  (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:6.
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
OS   Streptococcus pneumoniae.
PN   WO9818931-A2.
PD   07-MAY-1998.
PF   30-OCT-1997;  U19588.
PR   31-OCT-1996;  US-029960.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI   Kunsch CA, Rosen CA;
DR   WPI: 98-272225/24.
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
PS   Claim 1; Page 183-194; 1409pp; English.
CC   The present invention describes a computer readable medium which has
CC   recorded thereon sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC   on it, or a representative fragment or a sequence at least 95% identical
CC   to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC   to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC   pneumoniae. The present invention also describes an isolated nucleic acid
CC   molecule encoding a homologue of any of the fragments of the S.pneumoniae
```

```
18107 CGC.........GTGATGAAGATTCCATTATTACCA.........18133
          :::      |||   ||||||

  144 uArgValAsnGluProGluTrpLeuTrpProArgAsnProProIleArgG 161
                   ||||||||       :::::  ||||||   :
18134 .........GACCAGTGGCTAAAGGCGGAAAGTATGCCCATACTTC 18170

  161 luArgLysSerIleProThrSerTrp..........LeuLysIle 172
       ::  |||  :::||:::            :::  |||
18171 ATAC....AAGATTGTAGCTTCTTATGGAAATATTCACTTGGTCTATATT 18216

  173 ThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisVa 189
       |||:::  ||||||:::::||::||:::|||       ::||:::
18217 CACCTGCACACTGGTCGAACCCATCAAATCCGAGTCCATTTTCTCATAT 18266

  189 lGlyPheProThrLeuArgLeuIleArgTyrAlaMetGlySerTyrThrL 206
       :||||||||| |||  ::::::|||||: :::|||
18267 CGGTTTTCCTTTGCTGGGA..........................18285

  206 euAspSerLeuAlaAsnGlyGluTrpArgAsp 216
       |||:::|||  ::::::  :::|||
18286 ..GATGATTTGTATGGGTGGTAGTCTGGAAGAT 18315
```

OM of: US-09-252-691-7056 to: Issued_Patents_NA:*   out_format : pfs

Date: Jun 10, 2000 1:37 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL-frame+_p2n.model -DEV-xlp
-O=/cgn2_1/USPTO_spool/US09252691/runat_05062000_101650_615/app_query.fasta.1
-DB=Issued_Patents_NA -QMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: Issued_Patents_NA:*
Database sequences: 230463
Database length: 64992525
Search time (sec): 26.040000

score_list:

sequence	Strd	Orig	ZScore	EScore	Len	| Documentation
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-700-576-1	+	206.00	391.54	2.4e-14	3061	| Documentation
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-247-901C-1	-	94.50	126.08	14.84	50341	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-09-075-904-1	-	94.50	126.08	14.84	50341	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-07-952-817-8	-	86.00	142.14	1.89	2639	!
/cgn2_6/ptodata/1/1na/backfiles1.seq:5210025-1		86.00	142.14	1.89	2639	Pate
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-804-227C-1	+	85.50	108.98	132.97	43280	!
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-07-998-289B-7	+	83.50	128.56	10.80	5461	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-338-702-7	+	83.50	126.54	13.99	6513	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-337-339-7	+	83.50	126.54	13.99	6513	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-724-095-7	+	83.50	126.54	13.99	6513	!
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US95-14262-7	-	83.50	126.54	13.99	6513	!
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US95-14378-7	+	83.50	126.54	13.99	6513	!
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-781-802-7	-	79.50	126.52	14.03	3147	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-804-227C-7	+	79.50	96.14	690.37	44377	!
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-804-198-1	+	79.50	96.14	690.37	44377	!
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-438-439C-1	+	79.00	137.03	3.65	1150	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-705-245-19	+	79.00	137.03	3.65	1150	!
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-06665-1	+	79.00	137.03	3.65	1150	!
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-259-148A-9	-	79.00	135.52	4.42	1311	!
/cgn2_6/ptodata/1/1na/5B_COMB.seq:PCT-US95-13703-1	+	79.00	135.52	4.42	1311	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-07-876-941A-9	+	79.00	135.52	4.42	1311	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-194-981E-2	-	79.00	134.61	4.97	1419	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-07-876-941A-1	+	79.00	134.12	5.30	1482	!
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US95-13703-3	+	79.00	132.90	6.19	1647	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-240-049B-1	+	79.00	130.77	8.13	1983	!
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-259-148A-3	+	79.00	130.40	8.53	2049	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-484-054-3	+	79.00	130.40	8.53	2049	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-07-876-941A-3	-	79.00	130.40	8.53	2049	!
/cgn2_6/ptodata/1/1na/5B_COMB.seq:PCT-US95-13703-1	+	79.00	130.40	8.53	2049	!
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-259-148A-1	+	79.00	130.15	8.81	2094	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-484-054-1	+	79.00	130.15	8.81	2094	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-07-876-941A-1	-	79.00	130.15	8.81	2094	!
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-08849A-4	+	79.00	116.02	53.92	7168	!
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-08849-4	+	79.00	116.02	53.92	7168	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-808-793-24	+	77.50	113.97	70.12	6519	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-696-834-5	-	77.00	116.87	48.37	4624	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-125-468-1	-	77.00	95.40	758.86	30001	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-474-933-1	+	77.00	95.40	758.86	30001	!
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-651-136-5	-	76.50	134.27	5.19	927	!
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-635-121-1	+	76.50	114.62	64.52	5134	!
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-173-508-3	-	75.50	122.34	23.99	2185	!
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-265-310-3	-	75.50	122.34	23.99	2185	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-07-745-206A-6	+	75.50	110.93	103.66	5904	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-455-543A-3	+	75.50	110.93	103.66	5904	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-193-078B-3	+	75.50	110.93	103.66	5904	

seq_name: /cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-700-576-1

seq_documentation_block:
; Sequence 1, Application US/08700576
; Patent No. 5750848
; GENERAL INFORMATION:
; APPLICANT: Kruger, Niels
; APPLICANT: Steinbuchel, Alexander
; TITLE OF INVENTION: DNA SEQUENCE USEFUL FOR THE PRODUCTION
; TITLE OF INVENTION: OF POLY-3-HYDROXYALKANOATES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,576
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/787-1400
; TELEFAX: 713/789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-700-576-1

alignment_scores:
Quality: 206.00   Length: 189
Ratio: 1.856   Gaps: 6
Percent Similarity: 58.730   Percent Identity: 32.275

alignment_block:
US-09-252-691-7056 x US-08-700-576-1

Align seq 1/1 to: US-08-700-576-1   from: 1 to: 3061

39 GlnProThrArgValIleLeuPheAsnLysProTyrAspValLeuProG1 55
    ::|||:::||| :::::::||| |||||| :::
46 CGCCGGCCGCGTACCTGACCGCTACCTGGCTGGCTGGTGCGTAACGGC   95

55 nPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIleProVal. 71
    ||||||||||||||||||||| :::
96 CACCCACGATCCGACAACACGTACCGTTCTC...GACCTGCTGCCAGCGG 142

72 ....GlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGlu 85
            :::::: ||||||| :::::
143 CGTTGCGAGATGACCTGCACATAGCCGGGCGCCTGGACTTCAACACCACC 192

86 GlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGl 102
    ||||||||||||||||||| ::: ||| ||
-193 GGCCTGATGATCCTGACCAACGATGGCCAATGGTCACGGCGGCCTGACCA 242

```
102 nProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGluP 119
    ||||:::  :::  |||  |||  ||::::::|||  |||
243 GCCTGCCACCAAGCTGCCCAAGCATTATCTGGTGGACACCGAGGACGAGA 292

119 rOAspAspAlaSerLeuAlaLysLeuArgAsnGly......ValThrLeu 133
    :::: ||||||:::::|||||:|  :::||
293 TTGGCGGAGGCACTATGTGGCCAAGTTTCCGCAGGGTTTCTATTTTGCCTTC 342

134 AsnAspGlyProThrLeuProAlaGlyIleGluArgValAsnGluProGl 150
    :::||||  ||||||  :::::::
343 GAAGACCTCACCACCCAACCTGCCCAGCTGGACATCCTCGGC......... 384

150 uTrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProT 167
                   |||  |||  |||
385 ...............CCCCACCGAGCCCGG................... 399

167 hrSerTrpLeuLysIleThrLeuTyrGluLyrGlyArgAsnArgGlnValArg 183
    :::::::|||  ||||||||:::|||||||::
400 .............CTGGCGATCGTCGAGGGCGTTACCACCAGGTCAAG 435

184 ArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAl 200
    ||||||  :::  ||  ::  |||  |||  ::
436 CCCATGTTCGGGCATTTCAACACAAGGTGATCGGGCTGCATCGGGAGAG 485

200 aMetGlySerTyrThrLeuAsp...SerLeuAlaAsnGlyGluTrpArgA 216
    ::||||:::  |||||||  :::||||||  ||||||:::|||:
486 CATGGGGGCGATCCGGCTGGGATCCGGGGTTGGCGCCGGGAGTATCGTG 535

216 spValThrProLysGlu 221
    ::::|||  :::|||
536 AACTGACGGCCAATGAG 552
```

seq_name: /cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-247-901C-1

alignment_scores:
 Quality: 94.50   Length: 184
 Ratio: 1.243   Gaps: 8
Percent Similarity: 41.304   Percent Identity: 25.543

alignment_block:
US-09-252-691-7056 x US-08-247-901C-1/rev

Align seg 1/1 to reverse of: US-08-247-901C-1   from: 1 to: 50341

```
  62 ArgSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGl 78
     ||||:||:||  :::|||::::|||:::||:::::|||::|||||::| ||
6914 CGCCAGACGGTTCGGATCGTCATGCCCCCAGGGTATTCACGTACTCGG 6865

  78 yArgLeuAspArgAspSer.....GluGlyLeuLeu........ValL 90
     |::::     ||:||||||     ||||||||||||
6864 AGAGCCACTTCAGGACTCCCAGCCGAGGGTCTTCTCAGGCAGATACCAC 6815

  90 euThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArg 106
     |||||  |||  ::||||  :::|||  :::  |||  :::
6814 TCACCATCGACCGTCTTCTGCAGGACGGGCCCGATGATGTGAGGGGGAGA 6765

 107 ThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAspAlaSe 123
     |||||||:::   |||||   |||||||:::|||   |||
6764 CGGGGCAAGCTCCGGGTGGTTGTGGTGTTGAGGCCCACCCC......GCCTC 6721

 123 rLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuP 140
     |::::  :::::::   |||    :::||||  |||   |::|  |||
6720 CTTTCTCAGATTCGTCGTAGGA................ACTCGAC 6692

 140 roAlaGlyIleGluArgValAsnGluProGlu................ 150
     ||  |||    :::|||
6691 CGCCGGGAGCGGTTGTACAAATGCGGCCAGCTCGGCGCTCGGGCCTTGGTCGACT 6642
```

```
            STRAIN:
            INDIVIDUAL ISOLATE:
            DEVELOPMENTAL STAGE:
            HAPLOTYPE:
            TISSUE TYPE:
            CELL TYPE:
            CELL LINE:
            ORGANELLE:
        IMMEDIATE SOURCE:
            POSITION IN GENOME:
            CHROMOSOME/SEGMENT:
        FEATURE:
            NAME/KEY:
            LOCATION:
            IDENTIFICATION METHOD:
            OTHER INFORMATION:
        PUBLICATION INFORMATION:    No. 5994137e
            AUTHORS:
            TITLE:
            JOURNAL:
            VOLUME:
            PAGES:
            DATE:
            DOCUMENT NUMBER:
            FILING DATE:
            PUBLICATION DATE:
            RELEVANT RESIDUES IN SEQ ID NO:
US-09-075-904-1

alignment_scores:
        Quality:    94.50        Length:    184
        Ratio:    1.243        Gaps:    8
        Percent Similarity:  41.304    Percent Identity:  25.543

alignment_block:
US-09-252-691-7056 x US-09-075-904-1/rev
Align seg 1/1  to reverse of: US-09-075-904-1  from: 1  to: 50341

   62 ArgSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGl  78
      |||:|||:||   :||||||||||||||   :||||:|||||:||   ||
 6914 CGCCAGACGGTTCGGATCGTCATGCCCAGCCAGGGGATTCACGTACTACGG  6865

   78 yArgLeuAspArgAspSer....GluGlyLeuLeu.......ValL  90
      |:::|    :|||||||||||||||||||
 6864 AGAGCCACTTCAGGACTCCCCAGCCGAGGGTCTTCTCCAGGCAGATACCAC  6815

   90 euThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArg  106
      |:||||   |:||:|   |::||    |:||||
 6814 TCACCATCGACCGTCTTCTGCCAGGACGGGCCGATGATGTGAGGGGGAGA  6765

  107 ThrGlyLysIleTyrTyrValGlnValGluGlyValProAspAspAlaSe  123
      ||||||:||   |||||||:|:|||:||||
 6764 CGGGGGCAAGCTCCGGGGTGGTGGTTGTTGAGGCTCACCCC.....GCCTC  6721

  123 rLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuP  140
      |:||:   |||||||:||   |:||||
 6720 CTTTCTCAGATTCGTCGTAGGA.................ACTCGAC  6692

  140 roAlaGlyIleGluArgValAsnGluProGlu..................  150
      ||  |||    :|||:||
 6691 CGCCGGGAAGCGGTTGTACAAATGCGGCCACCTCGGGCCCTTGGGTCGACT  6642

  151 .:,.............................TrpLeuTrpPro........  154
            |||||||||||||
 6641 TGGCCGAAGAACGTGAGGGGCGTCGATGATGGCTCTGGCCATCCGGATTCCC  6592

  155 .......................................ArgAsnP  157
                      ||||:::||
 6591 TCCCAGATCGGACGCTGTCCAAGCTCGATGAGCTGGGCGATGACGGAGTC  6542

  157 ro.................................Pro 158
```

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-075-904-1

seq_documentation_block:
Sequence 1, Application US/09075904
Patent No. 5994137
(1) GENERAL INFORMATION:
    (i) APPLICANT: Jacobs, et al.
    (ii) TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
    (iii) NUMBER OF SEQUENCES: 1
    (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: Amster, Rothstein & Ebenstein
        (B) STREET: 90 Park Avenue
        (C) CITY: New York
        (D) STATE: New York
        (E) COUNTRY: U.S.A.
        (F) ZIP: 10016
    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
        (B) COMPUTER: IBM PC Compatible
        (C) OPERATING SYSTEM: MS-DOS
        (D) SOFTWARE: Word Processor (ASCII)
    (vi) CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER: US/09/075,904
        (B) FILING DATE: May 11, 1998
        (C) CLASSIFICATION:
    (vii) PRIOR APPLICATION DATA:
        (A) APPLICATION NUMBER: 08/247,901
        (B) FILING DATE: May 23, 1994
    (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: Bogoslan, Elizabeth A
        (B) REGISTRATION NUMBER: 39,911
        (C) REFERENCE/DOCKET NUMBER: 96700/475
    (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: (212) 697-5995
        (B) TELEFAX: (212) 286-0854 or 286-0082
        (C) TELEX: TWX 710-581-4766
(2) INFORMATION FOR SEQ ID NO: 1:
    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 50341
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear
    (ii) MOLECULE TYPE:
        (A) DESCRIPTION: L5 shuttle phasmid sequence
    (iii) HYPOTHETICAL: No
    (iv) ANTI-SENSE:
    (v) FRAGMENT TYPE:
    (vi) ORIGINAL SOURCE:
        (A) ORGANISM: L5 mycobacteriophage

```
  151 .................................TrpLeuTrpPro....  154
                              ||||||||||||
 6641 TGGCCGAAGAACGTGAGGGGCGTCGATGATGGCTCTGCCCATCCGGATTCCC  6592

  155 .........................................ArgAsnP  157
                          |||:::|
 6591 TCCCAGATCGGACGCGTGTCCAAGCTCGATGAGCTGGGCGATGAC  6542

  157 ro.................................Pro 158
            |||
 6541 CTTACCTCCGATGAATCGCCTAGCGCTCATCACGATTCGGCCAATGGCCA  6492

  159 IleArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTy  175
      |||   |||   |||                 ||   ::
 6491 CCTCGTACTGTGCATGTCGTCCTCCTTGATGGAGGCGTACATGTCGCCG  6442

  175 rGluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPheP  192
      ::::::||||:::||||      ||      ||||||::::    |
 6441 TCGTGCGCGTAGTCGGGACCTCAAAGCCGTACTGCTCGA...GGTTCTC  6395

  192 ro 192
     ||
 6394 CA 6393
```

```
6541 CTTACCTCCGATGAATCGCCTAGCGCTCATCACGATTCGGCCAATGGCCA 6492
              |||
 159 IleArgGluArgLysSerIleProThrSerPheArgLysIleThrLeuTyr 175
              |||          |||             :::
6491 CCTCGTACTCGTGCATGTCGTCCTCCTTGATGAGGCGGTACATGTCGGCCG 6442

 175 rGluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPheP 192
      :::::::|||::||||     ::       ||||||:::    |||  -
6441 TCGTGCGCGTAGTCGCGGACCTCAAAGCCGTACTGCTCGA...GGTTCTC 6395

 192 ro 192
      ||
6394 CA 6393
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-952-817-8

seq_documentation_block:
; Sequence 8, Application US/07952817
; Patent No. 5356796
; GENERAL INFORMATION:
;   APPLICANT: Keller, John W.
;   TITLE OF INVENTION: A Repressor Protein and Gene for Regulating
;   TITLE OF INVENTION: Expression of Polypeptides and Its Use in the Preparation of
;   TITLE OF INVENTION: 2,2-Dialkylglycine Decarboxylase of Pseudomonas Cepacia
;   NUMBER OF SEQUENCES: 30
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W., Suite 700
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: US
;   ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/952,817
;   FILING DATE: 19920928
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Meyers, Kenneth J.
;   REGISTRATION NUMBER: 25,146
;   REFERENCE/DOCKET NUMBER: 01120.0002-01000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-408-4000
;   TELEFAX: 202-408-4400
;   INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2639 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-07-952-817-8

alignment_scores:
        Quality: 86.00   Length: 221
          Ratio: 0.789   Gaps: 7
Percent Similarity: 49.321   Percent Identity: 24.434

alignment_block:
US-09-252-691-7056 x US-07-952-817-8/rev

Align seg 1/1 to reverse of: US-07-952-817-8 from: 1 to: 2639

   11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSer 27
      :::::||:::||| |||||| |:||| |:|||
 1991 TCGTCAATACGCAGGCTTCGGCGCGGCCGCCGCCG.......AG 1951
```

```
   27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
      |||||||:::::::|||||||     :::            ::|||||||
 1950 CTCACGCCGCTCGACACCGTGCGTGCGTGTTCGAAACCAACACGATCGG 1901

   44 leLeuPheAsnLysProTyrAspValLeuProGlnPhe............. 56
      :::::             |||||||||||||||
 1900 CACGATTCGGTGACGCAGGGCCCGTGCCCAGTTTCGCCGGCGTGG 1851

   57 ..........ThrAspGluAlaGlyArgSerThrLeuLysAspPheI 69
                :::::       |||||||||        :::
 1850 GCCGGCGTGATCGTGAACGTCACGTCGAGCGTCACGCTGAAG...GTGTT 1804

   69 eProValGlnGlyValTyrAlaAlaGlyArg.....LeuAspArgAspS 84
      :|||:::    |||:::|||     ||||||         :::
 1803 GCCCGTGGTCGGCGCCATACCGGGCCAGCAAGGCGCGGTCAATGCGTTCA 1754

   84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
      :||:::||  ||     |||                 ||:::|||:::||| •
 1753 CCGAATCGATGGCGGTCGAACTGAACCGTTCGGCGTGCGCGCCATCTC 1704

  101 ThrGlnProGlyLys..................................... 105
      :::  ||||||:::
 1703 GTGCTGCCGGGCCGTGCACCCGACACGCGTTTCGGCGACAACGCGCGTG 1654

  106 ...........ArgThrGlyLysIleTyrTyrValGlnValGlyGluG 117
                 |||||       ::|||     :::|||
 1653 CCGCACATCATGGCTTCGAACACGAGGCTTATGCGAGTTCGCCGGAAAG 1604

  117 lyGluProAspAspAlaSerLeuAlaLys....................... 126
      |||:::       :::                ::
 1603 GCCGTCGCGAGGATGCTCGATGCGTCGGGCGGCGATCACCCGTGCGCAGGA 1554

  127 ...LeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGl 142
         |||     :::  :::  |||::::|||||:::          ||:::
 1553 TGTTGCCGACGCCGTGTGGCGCGCGGACCCGTCGAGCCCGATGC 1504

  142 yIleGluArgValAsnGluProGluTrpLeuTrpProArgAsnProProI 159
         :::  |||||             :::   ::|||       ||||||
 1503 GCATTCGCCGGGTGCCGATGCGAAGCGTGGCGCGGCCGAAACGCGCTGA 1454

  159 leArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyr 175
      |||:::||:::::::     :::      :::|||      |||:::
 1453 TCGATCGGAATCCGTGCGCGCGGTCCTCGAACCGCGCGACG...... 1410

  176 GluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhePr 192
      |||     ||:::|||::: |||||||  ::::|||     ::|||
 1409 AATCGCACGATCGGAGAAGGGCTTACCGGCGGCGGCGCGT 1369

  192 oThrLeuArgLeu 196
      |||  |||:::
 1368 TACAGCGCCGGTT 1356
```

seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5210025-1

seq_documentation_block:
;Patent No. 5210025
;   APPLICANT: KELLER, JOHN W.
;   TITLE OF INVENTION: REPRESSOR PROTEIN GENE FOR REGULATING
;EXPRESSION OF POLYPEPTIDES AND ITS USE IN THE PREPARATION OF
;2,2-DIALKYLGLYCINE DECARBOXYLASE OF PSEUDOMONAS CEPACIA
;   NUMBER OF SEQUENCES: 18
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/501,814
;   FILING DATE: 30-MAR-1990
;SEQ ID NO:1:
;   LENGTH: 2639
5210025-1

alignment_scores:
        Quality: 86.00   Length: 221
          Ratio: 0.789   Gaps: 7

Tue Jun 13 10:30:13 2000

Percent Similarity: 49.321   Percent Identity: 24.434

alignment_block:
US-09-252-691-7056 x 5210025-1/rev

Align seg 1/1 to reverse of: 5210025-1   from: 1 to: 2639

```
   11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
      ::|||||||||:|||::|||  |||  ::|||     ||
 1991 TCGTCAATAACGCAGGCTTCGGCGCGGCCGCCGCCG........AG 1951

   27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
      :|||||||||::|||| .:|:    ::|||||||||
 1950 CTCACGCGCCTCGACACGGTGCGTGCCGTGTTCGAAACCAACACGATCGG 1901

   44 leLeuPheAsnLysProTyrAspValLeuProGlnPhe.......... 56
      :::::::||||||||||||
 1900 CACGATTGCGGTGACGCAGGGCCGTGCTGCCCAGTTCGCCGGGGTGGG 1851

   57 ..........ThrAspGluAlaGlyArgSerThrLeuLysAspPheIl 69
                :::::|| ||||||||
 1850 GCCGGCGTGATCGTCAACGTCACGTCGAGCGTCACGCTGAAG...GTGTT 1804

   69 eProValGlnGlyValTyrAlaAlaGlyArg......LeuAspArgAspS 84
      ::|:::: ||||:||| :::: ||
 1803 GCCGCTGGTCGGCGCATACCGGGCCAGCAAGGCGGCGGTCAATGCGTTCA 1754

   84 erGluGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeu 100
      ::||:.::|||| ||| :::|:|::|||::|||
 1753 CCGAATCGATGGCGGTCGAACTCGAACCGTTCGGCGTGCCGCGCATCTC 1704

  101 ThrGlnProGlyLys.................................. 105
      :::  ||||||:::
 1703 GTGCTGCCGGGCCGTGCACCCGACCGCGTTCGGCGACAACGCGCGTG 1654

  106 ..........ArgThrGlyLysIleTyrTyrValGlnValGluG 117
                ::|||  ||||:: :::|
 1653 CGCACATGCATGGCTTCGAACACGAGGCTTATGCGGAGTTCGCCGGAAAG 1604

  117 lyGluProAspAspAlaSerIeuAlaLys.................... 126
      |:|:: .::||||||
 1603 GCCGTCGCGGAGGATGCTCGATGCGTCGGGGGCGATCACCCGTGCCAGGA 1554

  127 ..LeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGl 142
      ||||::  |||::: .:|||||||:::: |||  :|
 1553 TGTTGCCGACGCCGTGTGGGCGGCGACCCGTGCCGAGCCCGATGC 1504

  142 yIleGluArgValAlaAsnGluProGluTrpProLeuTrpProArgAsnProProI 159
      :  |||   ||| :::
 1503 GCATTCCGGCGCCGGTGCCGATGCCGAAGCGTGGGCGGCCGAAACGCGCTGA 1454

  159 leArgGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyr 175
      ||     |:::|||::: ||:::::::
 1453 TCGATCGGGAATCCGTCCGCGCCGGTCGCTCGAACCGGCGACG....... 1410

  176 GluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGlyPhePr 192
      ||||||  ::: ::|||||    |||
 1409 ....AATCCGCACGATGCGGAGAAGGGCGTTACCGGCGGCGGCGGCGT 1369

  192 oThrLeuArgLeu 196
      |||  |: :|:
 1368 TACAGCGCGCGTT 1356
```

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
  ADDRESSEE: THOMAS G. PLANT 1501
  STREET: LILLY CORPORATE CENTER
  CITY: INDIANAPOLIS
  STATE: IN
  COUNTRY: USA
  ZIP: 46285
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: MS-DOS
  SOFTWARE: ASCI(DOS) Text only
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/804,227C
  FILING DATE: February 21, 1997
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: Plant, Thomas, G.
  REGISTRATION NUMBER: 35,784
  REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 43280 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
  NAME/KEY: CDS
  LOCATION: 816..14234
FEATURE:
  NAME/KEY: CDS
  LOCATION: 14351..19945
FEATURE:
  NAME/KEY: CDS
  LOCATION: 20010..31199
FEATURE:
  NAME/KEY: CDS
  LOCATION: 31232..36067
FEATURE:
  NAME/KEY: CDS
  LOCATION: 36249..41774
US-08-804-227C-1

alignment_scores:
    Quality: 85.50   Length: 202
    Ratio: 0.891   Gaps: 9
Percent Similarity: 47.525   Percent Identity: 25.248

alignment_block:
US-09-252-691-7056 x US-08-804-227C-1

Align seg 1/1 to: US-08-804-227C-1 from: 1 to: 43280

```
   19 ArgLysHisArgValGluArgPheSerSerArgGlnAlaThrArgArgTh 35
      :::||||::|||  ::::||| .::||||  ::|||||
41806 CGGCGGCACCACGCCCCCGCCGGCGGCGGCACACCTGTCCCGCCCCGGAC 41855

   35 rProGluPro...GlnProThrArgValIleLeuPheAsnLysProTyrA 51
      | :|||| :::|| :::||||  ||::::
41856 GCAGGCTCCAACTCACCCGGCGCCGCCAGTGGTTCGCCGGCAACCA.... 41901

   51 spValLeuProGlnPheThrAspGluGlyArg..........Ser 63
      ||   :::
41902 ..GGGAGACCCCTACGGATGATCCTGCCGGCCGCGGCCGACCCGGC 41949

   64 ThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGlyArgLe 80
      ::|| ||| :::::: :::||
```

Tue Jun 13 10:30:13 2000

```
41950 ACCGTACCGAGGAAGAGATCCGTGAGCGGGGCCGCTGTTCCAGCGAAC 41999
   80 uAspArg.........AspSerGluGlyLeuLeuValLeuThrAsnAspG  94
          ||         ::||||
42000 TCCTCGGCCGCTCGGTGACCGGCAGCCCCATGTCGCCGACGCCGTGACG 42049
   94 lyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLys.Il 110
       ||       :::||||||| :||:||||| :: ||
42050 GCCGACGACGCGTTCGGCCGCCCTCACCGCGACGGTGCACGGCCAGGAGT 42099
  110 eTyrTyrTyrValGlnValGluGlyGluProAspAspAlaSerLeuAlaLysL 127
          : :: ::||||||||:: :|||:::
42100 CCGGAACTGCCCCTCTCCGGCAGCGCCCTCGACGCCGCCCACGGGAAC. 42148
  127 euArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGlyIle 143
          ||||||||: |||||:::||
42149 ..........CCCGGCCGCCCGCCCCTCCCCGGAGGG... 42175
  144 GluArgArgValAsnGluProGlurpLeuTrpProArgAsnProIleAr 160
          ||||||||:::::||||| ||
42176 ................TGGCCGCACCGGCCCCCCGGACAG 42198
  160 gGluArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrGluG 177
       ||||:::||: |||       |||       :
42199 GGAGGAGCGAGACGACCCCGACCCGCACCGGCGGACCTGCTGAACGCCG 42248
  177 lyArgAsnArgGlnVal............. ArgArgMetThr 186
       ::           ||||||       ||||||:||:||:
42249 CCCGGCCCGGGGCAGGTCCTCGACCTCGTCCCGTTCGCCCGGCGCTGGCG 42298
  187 AlaHis.............ValGlyPheProThrLeuArgLeuIleAr 198
       ||:::           :::||||       |||::: ||
42299 GCCCGGACGGCCGGCGCGTGGCGTCCCGGCGTCCCGGCGACGGCTGCCCG 42348
  198 gTyr 199
       |:::
42349 CTTC 42352
```

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-07-998-289B-7

```
alignment_scores:
    Quality:      83.50      Length:   133
    Ratio:         1.265       Gaps:     6
Percent Similarity: 49.624   Percent identity: 30.075

alignment_block:
US-09-252-691-7056 x US-07-998-289B-7

Align seg 1/1   to: US-07-998-289B-7   from: 1   to: 5461

   11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe  27
      ::||||||||||||| |||       :::||| |||
 1936 AGACAATGACCAAGGAGGAGCAAATTGCGCAACCGCAACACACGCAATCA 1985

   27 rSerArgGlnAlaThrArgArgThrProGlnProGlnProThrArgVall  44
      :||| |||||||       ||||||::|||
 1986 ATCAGTGGGCGGCCACCATGGGCGACCACCTGTCTGACACCAATCACA   2035

   44 leLeuPheAsnLysProTyrArgValLeuProGlnPheThrAspGluAla  60
      ||| ::::||:||||       :::       ||||||||||||||||
 2036 AGCTCGATCATCCGCACTACGAAATTGGCCTGGAGTGCACGGACGAAGCT 2085

   61 GlyArgSerThrLeuLysAsp......PheIle...ProValGlnGlyVva  74
      |||:::                     ||||||       ||||||||||
 2086 GGCAAGATTAAACATCATGACAATCCTTTATCGAGCCCGTCCAGACACA  2135

   74 lTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuT  91
         :::||       |||       ||  |||:::||||||
 2136 AACGGTGGTT............GATATGAAAGATGTGATGGTCCTG.    2169

   91 hrAsnAspGlyValLeuGlnAla.....ArgLeuThrGlnProGlyLys  105
      ||||||       |:: |||||       ||       ||  :::||||
 2170 ..AATGACAATCATCGAACACGGCCGCTGGTCGGCACAGTCGGGCAAGCGAT 2217

  106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAspAl  122
      ||       ::::||||||||       :::||
 2218 CGGGGTGTCTCCGTTTACTATTTCCCAACAGAGGACGATGACGAG.....  2262

  122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr  138
                                      |||||||||||
 2263 ...............................GATGGGCCGACG    2274
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-338-702-7

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/338,702
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wallen III, John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19338
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6513 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-338-702-7

alignment_scores:
Quality: 83.50   Length: 133
Ratio: 1.265   Gaps: 6
Percent Similarity: 49.624   Percent Identity: 30.075

alignment_block:
US-09-252-691-7056 x US-08-338-702-7

Align seg 1/1 to: US-08-338-702-7 from: 1 to: 6513

    11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
       :::||||||||||| ||| :::::||| ||| :: :
  2022 AGCACAATGACCAAGGAGGAGCAAATTGGCAACCGCAACCACACGCCAATCA 2071

    27 rSerArgGlnAlaThrArgArgThrProGlnProGlnProThrArgValI 44
       ||:||||:
  2072 ATCAGTGGCGGCCACCAATGGCGGCCACCACCTGTCTGGACACCAATCACA 2121

    44 leLeuPheAsnLysProTyrValGluProGlnPheThrAspGluAla 60
       |||| :::::||||||||||
  2122 AGCTCGATCATCGCGACTACGAAATTGGCCTGGAGTGCACGGACGAAGCT 2171

    61 GlyArgSerThrLeuLysAsp......Phelle...ProValGlnGlyIVa 74
       ||:::||||| ||||||| ||||||||
  2172 GGCAAGATTAAACATCATGACAATCCTTTATCGAGCCGTCCAGACACA 2221

    74 lTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuT 91
       ::::::||||| ||| :: :
  2222 AACGGTGGTT............GATATGAAAGATGTGATGGTCCTG. 2255

    91 hrAsnAspGlyValLeuGlnAla......ArgLeuThrGlnProGlyLys 105
       |||||| ::: ||||| |||
  2256 ..AATGACATCATCGAACAGACGCCGCTGGTCGGCACAGTCGGGCAAGCGAT 2303

seq_name: /cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-337-339-7

seq_documentation_block:
; Sequence 7, Application US/08337339
; Patent No. 5593864
; GENERAL INFORMATION:
APPLICANT: Warmke, Jeffrey W.

APPLICANT: Hall, Linda
APPLICANT: Feng, Gouping
APPLICANT: Van Der Ploeg, Leonardus
TITLE OF INVENTION: PROCESS FOR FUNCTIONAL EXPRESSION OF THE
TITLE OF INVENTION: PARA SODIUM CHANNEL
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: John W. Wallen III
STREET: P.O. Box 2000, 126 E. Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/337,339
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wallen III, John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19332
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6513 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-337-339-7

alignment_scores:
Quality: 83.50   Length: 133
Ratio: 1.265   Gaps: 6
Percent Similarity: 49.624   Percent Identity: 30.075

alignment_block:
US-09-252-691-7056 x US-08-337-339-7

Align seg 1/1 to: US-08-337-339-7 from: 1 to: 6513

    11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
       ::||||||||||||| ||| :::::||| ||| :: :
  2022 AGCACAATGACCAAGGAGGAGCAAATTGGCAACCGCAACCACACGCCAATCA 2071

    27 rSerArgGlnAlaThrArgArgThrProGlnProGlnProThrArgValI 44
       ||:||||:
  2072 ATCAGTGGCGGCCACCAATGGCGGCCACCACCTGTCTGGACACCAATCACA 2121

    44 leLeuPheAsnLysProTyrValGluProGlnPheThrAspGluAla 60
       |||| :::::||||||||||
  2122 AGCTCGATCATCGCGACTACGAAATTGGCCTGGAGTGCACGGACGAAGCT 2171

    61 GlyArgSerThrLeuLysAsp......Phelle...ProValGlnGlyIVa 74
       ||:::||||| ||||||| ||||||||
  2172 GGCAAGATTAAACATCATGACAATCCTTTATCGAGCCGTCCAGACACA 2221

    74 lTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuT 91
       ::::::||||| ||| :: :
  2222 AACGGTGGTT............GATATGAAAGATGTGATGGTCCTG. 2255

    91 hrAsnAspGlyValLeuGlnAla......ArgLeuThrGlnProGlyLys 105
       |||||| ::: ||||| |||
  2256 ..AATGACATCATCGAACAGACGCCGCTGGTCGGCACAGTCGGGCAAGCGAT 2303

```
106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAspAl 122
        |||  ||||||||||   :::|||   ::  |||
2304 CGCGGTGTCTCCGTTTACTATTTCCCAACAGAGGACGATGACGAG...... 2348
122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
                                        ||||||||||||
2349 ......................................GATGGGCCGACG 2360

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-724-095-7

seq_documentation_block:
; Sequence 7, Application US/08724095
  Patent No. 5688917
; GENERAL INFORMATION:
;   APPLICANT: Warmke, Jeffrey W.
;   APPLICANT: Hall, Linda
;   APPLICANT: Feng, Gouping
;   APPLICANT: Van Der Ploeg, Leonardus
;   TITLE OF INVENTION: PROCESS FOR FUNCTIONAL EXPRESSION OF THE
;   TITLE OF INVENTION: PARA SODIUM CHANNEL
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: J. Mark Hand - Merck & Co., Inc.
;     STREET: P.O. Box 2000 - 126 E. Lincoln Avenue
;     CITY: Rahway
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07065-0907
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: FASTseq. Version #1.d5
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/724,095
;     FILING DATE:
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Hand, J. Mark
;     REGISTRATION NUMBER: 36,545
;     REFERENCE/DOCKET NUMBER: 19332DA
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (908) 594-3905
;     TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 6513 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
US-08-724-095-7

alignment_scores:
        Quality: 83.50          Length: 133
        Ratio:   1.265          Gaps:   6
Percent Similarity: 49.624    Percent Identity: 30.075

alignment_block:
US-09-252-691-7056 x US-08-724-095-7

Align seg 1/1  to: US-08-724-095-7  from: 1 to: 6513

11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe 27
   ::||||||||||||   |||       ::::|||     ||       ::
2022 AGCAACAATGACCAAGGAGGAGCAAATTGCCCAACCGCAACACACGCAATCA 2071
27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI 44
   |||:::|               |||:|:: 
2072 ATCAGTGGGCGCCACCAATGGCGCCACCACCTGTCTGGACACCAATCACA 2121
44 leLeuPheAsnLysProTyrAspValLeuProGlnPheThrAspGluAla 60
```

```
2122 AGCTCGATCATCGCGACTACGAAATTGGCCTGGAGTGCACGGACGGAAGCT 2171
                                     ::       |||||||||||
61 GlyArgSerThrLeuLysAsp......Phe...Ile...ProValGlnGlyVa 74
   ||:::                              ::||||||
2172 GGCAAGATTAAACATCATGACAATCCTTTATCGAGCCCGTCCAGACACA 2221
74 lTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuT 91
           |||          :::  |||||:||   ::|||||
2222 AACGGTGGTT................GATATGAAAGATGTGATGGTCCTG. 2255
91 hrAsnAspGlyValLeuGlnAla......ArgLeuThrGlnProGlyLys 105
   |||||| :::|||||  |||      |||    ::::  :::
2256 ..AATGACATCATCGAACAGGCCGTGGTCGCCACAGTCGGGCAAGCGAT 2303
106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAspAl 122
    |||   ||||||||||   :::|||   ::  |||
2304 CGCGGTGTCTCCGTTTACTATTTCCCAACAGAGGACGATGACGAG...... 2348
122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr 138
                                        ||||||||||||
2349 ......................................GATGGGCCGACG 2360

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-14262-7

seq_documentation_block:
; Sequence 7, Application PC/TUS9514262
; GENERAL INFORMATION:
;   APPLICANT: Warmke, Jeffrey W.
;   APPLICANT: Van Der Ploeg, Leonardus
;   TITLE OF INVENTION: PROCESS FOR FUNCTIONAL EXPRESSION OF THE
;   TITLE OF INVENTION: PARA SODIUM CHANNEL
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Jack L. Tribble
;     STREET: P.O. Box 2000, 126 E. Lincoln Avenue
;     CITY: Rahway
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07065-0907
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US95/14262
;     FILING DATE:
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Tribble, Jack L.
;     REGISTRATION NUMBER: 32,633
;     REFERENCE/DOCKET NUMBER: 19338 PCT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (908) 594-5321
;     TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 6513 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
PCT-US95-14262-7

alignment_scores:
        Quality: 83.50          Length: 133
        Ratio:   1.265          Gaps:   6
Percent Similarity: 49.624    Percent Identity: 30.075

alignment_block:
.US-09-252-691-7056 x PCT-US95-14262-7
```

Right column

```
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
PCT-US95-14378-7

alignment_scores:
        Quality:  83.50              Length:     133
        Ratio:    1.265              Gaps:         6
Percent Similarity:  49.624      Percent Identity:  30.075

alignment_block:
US-09-252-691-7056 x PCT-US95-14378-7

Align seg 1/1  to: PCT-US95-14378-7  from:  1  to:  6513

     11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe  27
        :::||||||||||| |||      :::::::||| ::          ::
   2022 AGCACAATGACCAAGGAGGAGCAAATTGCCAACCGCAACACCGCAATCA 2071

     27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI  44
        :|||     ||||||::   ||||||||||||||||
   2072 ATCAGTGGGCGCCACCAATGGCGGCACCACCTGTCTGGACACCAATCACA 2121

     44 leLeuPheAsnLysProTyrAspValLeuProGlnPheThrArgGluAla  60
        |||  ::::::   ||  |||||   ||||||||||
   2122 AGCTCGATCATCGCGACTACGAAATTGGCCTGGAGTGCACGGACGAAGCT 2171

     61 GlyArgSerThrLeuLysAsp......PheIle...ProValGlnGlyVa  74
        ||:::       |||||| ||||||||
   2172 GGCAAGATAAAACATCAAGACAATCCTTTTATCGAGCCCGTCCAGACACA 2221

     74 lTyrAlaAlaGlyArgLeuAspArgAspSer.....PheIle...ProValGlnGlyVa  74
        :::::    |||  :::|||||||
   2222 AACGGTGGTT..............GATATGAAAGATGTGATGGTCCTG.  2255

     91 hrAsnAspGlyValLeuGlnAla......ArgLeuThrGlnProGlyLys 105
        ||||||  :::   ||||||   |||    :::::: ::
   2256 ..AATGACATCGAACAGCCCGTGGTCGGCACGTCGGCCAAGCGAT      2303

    106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyProAspAspAspAl 122
        |||      ||||||||||:::||:::::::||||||
   2304 CGCGGTGTCTCCGTTTACTATTTCCCAACAGAGGACGATGACGAG...... 2348

    122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr  138
        |||           |||||||||||
   2349 ........................................GATGGGCCGACG 2360

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-781-802-7
```

Left column

```
Align seg 1/1  to: PCT-US95-14262-7  from:  1  to:  6513

     11 AsnThrMetThrLysThrSerPheArgLysHisArgValGluArgPheSe  27
        :::||||||||||| |||      :: ::
   2022 AGCACAATGACCAAGGAGGAGCAAATTGCCAACCGCAACACCGCAATCA 2071

     27 rSerArgGlnAlaThrArgArgThrProGluProGlnProThrArgValI  44
        :|||     ||||||::   ||||||||||||||||
   2072 ATCAGTGGGCGCCACCAATGGCGGCACCACCTGTCTGGACACCAATCACA 2121

     44 leLeuPheAsnLysProTyrAspValLeuProGlnPheThrArgGluAla  60
        |||  ::::::   ||  |||||   ||||||||||
   2122 AGCTCGATCATCGCGACTACGAAATTGGCCTGGAGTGCACGGACGAAGCT 2171

     61 GlyArgSerThrLeuLysAsp......PheIle...ProValGlnGlyVa  74
        ||:::       |||||| ||||||||
   2172 GGCAAGATAAAACATCAAGACAATCCTTTTATCGAGCCCGTCCAGACACA 2221

     74 lTyrAlaAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuT  91
        :::::    |||  :::|||||||
   2222 AACGGTGGTT..............GATATGAAAGATGTGATGGTCCTG.  2255

     91 hrAsnAspGlyValLeuGlnAla......ArgLeuThrGlnProGlyLys 105
        ||||||  :::   ||||||   |||    :::
   2256 ..AATGACATCGAACAGCCCGTGGTCGGCACGTCGGCCAAGCGAT      2303

    106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyProAspAspAspAl 122
        |||      ||||||||||:::||:::::::||||||
   2304 CGCGGTGTCTCCGTTTACTATTTCCCAACAGAGGACGATGACGAG...... 2348

    122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThr  138
        |||           |||||||||||
   2349 ........................................GATGGGCCGACG 2360

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-14378-7
```

```
          OPERATING SYSTEM: PC-DOS/MS-DOS
          SOFTWARE: PatentIn Release #1.0, Version #1.30
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/781,802
          FILING DATE: 10-JAN-1997
          CLASSIFICATION: 536
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 08/694,078
          FILING DATE: 07-AUG-1996
          CLASSIFICATION: 536
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 60/019,580
          FILING DATE: 12-JUN-1996
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 60/009,704
          FILING DATE: 11-JAN-1996
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 60/001,995
          FILING DATE: 01-AUG-1995
     ATTORNEY/AGENT INFORMATION:
          NAME: Chao, Mark
          REGISTRATION NUMBER: 37,293
          REFERENCE/DOCKET NUMBER: 95,963-E
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: 312-913-0001
          TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 7:
     SEQUENCE CHARACTERISTICS:
          LENGTH: 3147 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
     MOLECULE TYPE: cDNA
     FEATURE:
          NAME/KEY: CDS
          LOCATION: 245..1231
          OTHER INFORMATION: /note= "TspA E101 sequence longest
          OTHER INFORMATION: open reading frame; other possible start codons are TTG/leu9;
          OTHER INFORMATION: TTG/leu13; TTG/leu15; GTG/val43"
     FEATURE:
          NAME/KEY: mat_peptide
          LOCATION: 245..1231
US-08-781-802-7
```

```
alignment_scores:
          Quality:    79.50              Length:    128
          Ratio:      1.242              Gaps:        9
Percent Similarity:  50.000       Percent Identity:  31.250

alignment_block:
US-09-252-691-7056 x US-08-781-802-7/rev
Align seg 1/1  to reverse of: US-08-781-802-7  from: 1  to: 3147

  31 AlaThrArgArgThrProGlnProGluProThrArgValIleLeuPheAs 47
     |||  ||| |||::::||::||         .:::
1436 GCCCGGCTAAGACCCCCAACCCCCGCCTCAGC.............. 1404

  47 nLysProTyrAspValLeuProGlnPheThrAspGluAlaGlyArgSerT 64
     ||||||| |||   |||::::
1403 ...GCCTATGACCGGCTC.............CGCACCA 1382

  64 hrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaAlaGlyArgLeu 80
     ::||| |||   :::  ||||||||:|||         :
1381 GCCTGGCGGACGCCCTCTACTTCCAAGGGGTCTTCGCC..........CTG 1341

  81 AspArgAspSerGluGlyValLeuLeuValLeuThrAsnAspGlyValLeuGl 97
     |||      ::    |||||||:||||||||||
1340 AAGGCCGAC.........GTCCTGGTGGCGAAAGGGGAGGGGTCCTCCG 1300

  97 nAlaArgLeuThrGlnProGlyLysArgThrGlyLysIleIleLysTyrTyrValG 114
     :||:  |||         :::  |||   |||
```

```
1299 GGCCCTCCTCCAGGCCCCTCCCCAAG.......GAGGCGTACCGGGTCC 1256
 114 lnValGluGlyGluProAspAspAlaSerLeuAlaLysLeuArgAsnGly 130
     :::||||::: |||||   |||:::  ||::||:::||||
1255 TCCTCGAGGCCCTCCCCGAC.......CTCACGGACCTACGCGGGGGG 1215
 131 ValThrLeuAsnAspGlyProThrLeuProAlaGlyIleGluArgValAs 147
     :::||||::: |||   ||||:::::
1214 ..........CCTGCCCTTCTCCGCCAGCTCCTGAGCC......... 1185
 147 nGluProGluTrpProProArgAsnProPro 158
     ::||||||   |||::||||||
1184 .AGTCCAGAA..........CGCTCACCTCCC 1164
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-7

seq_documentation_block:
; Sequence 7, Application US/08804227C
Patent No. 5876991
; GENERAL INFORMATION:
     APPLICANT: DeHoff, Bradley S.
     APPLICANT: Kuhstoss, Stuart A.
     APPLICANT: Rosteck, Paul R., Jr.
     APPLICANT: Sutton, Kimberly L.
     TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
     NUMBER OF SEQUENCES: 15
     CORRESPONDENCE ADDRESS:
          ADDRESSEE: THOMAS G. PLANT 1501
          STREET: LILLY CORPORATE CENTER
          CITY: INDIANAPOLIS
          STATE: IN
          COUNTRY: USA
          ZIP: 46285
     COMPUTER READABLE FORM:
          MEDIUM TYPE: Floppy disk
          COMPUTER: IBM Compatible
          OPERATING SYSTEM: MS-DOS
          SOFTWARE: ASCI(DOS) Text only
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/804,227C
          FILING DATE: February 21, 1997
          CLASSIFICATION: 435
     ATTORNEY/AGENT INFORMATION:
          NAME: Plant, Thomas, G.
          REGISTRATION NUMBER: 35,784
          REFERENCE/DOCKET NUMBER: X-8231
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
     SEQUENCE CHARACTERISTICS:
          LENGTH: 44377 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
     MOLECULE TYPE:  DNA (genomic)
     FEATURE:
          NAME/KEY: CDS
          LOCATION: 350..14002
     FEATURE:
          NAME/KEY: CDS
          LOCATION: 14046..20036
     FEATURE:
          NAME/KEY: CDS
          LOCATION: 20110..31284
     FEATURE:
          NAME/KEY: CDS
          LOCATION: 31329..36071
     FEATURE:
          NAME/KEY: CDS
          LOCATION: 36155..41830
US-08-804-227C-7
```

```
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
    ADDRESSEE: PAUL R. CANTRELL, 1138
    STREET: LILLY CORPORATE CENTER
    CITY: INDIANAPOLIS
    STATE: IN
    COUNTRY: USA
    ZIP: 46285
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: Macintosh
    OPERATING SYSTEM: Macintosh 7.0
    SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/804,198
    FILING DATE:
    CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
    NAME: CANTRELL, PAUL R.
    REGISTRATION NUMBER: 36,470
    REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
    LENGTH: 44377 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
    NAME/KEY: CDS
    LOCATION: 350..14002
FEATURE:
    NAME/KEY: CDS
    LOCATION: 14046..20036
FEATURE:
    NAME/KEY: CDS
    LOCATION: 20110..31284
FEATURE:
    NAME/KEY: CDS
    LOCATION: 31329..36071
FEATURE:
    NAME/KEY: CDS
    LOCATION: 36155..41830
US-08-804-198-1
```

```
alignment_scores:
    Quality:       79.50          Length:     215
    Ratio:         0.914          Gaps:       10
Percent Similarity:  40.465    Percent Identity: 24.186

alignment_block:
US-09-252-691-7056 x US-08-804-198-1

Align seg 1/1  to: US-08-804-198-1  from: 1  to: 44377

    3 MetArgGlnLeuIleLeuThrProGluAsnThrMetThrLysThrSerPheAr 19
        :::||||||  ::::||||| :::::|
33596 CTTCGCCGGACACTCACCCCAGGTGGACA..................... 33625

   19 gLysHisArgValGluArgPheSerSerArgGlnAlaThrArgArgThrP 36
        |||    |||   ::: ||:||||    :::|
33626 .........GTTCCGCGCCGAGCTCCTCGACACCGTGGGCACCGTCCGC 33665

   36 roGluProGlnProThrArgValIleLeuPheAsnLysProTyrAspVal 52
        ||:||  ::::|||
33666 CCGACCGCCGCCCG.............................. 33679

   53 LeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIl 69
                                               33679
```

```
alignment_scores:
    Quality:       79.50          Length:     215
    Ratio:         0.914          Gaps:       10
Percent Similarity:  40.465    Percent Identity: 24.186

alignment_block:
US-09-252-691-7056 x US-08-804-227C-7

Align seg 1/1  to: US-08-804-227C-7  from: 1  to: 44377

    3 MetArgGlnLeuIleThrProGluAsnThrMetThrLysThrSerPheAr 19
        ::::|||||  ::::||||||| :::::|
33596 CTTCGCCGGACACTCACCCCAGGTGGAGCA.................... 33625

   19 gLysHisArgValGluArgPheSerSerArgGlnAlaThrArgArgThrP 36
        |||    |||   ::: ||:||||    :::|||
33626 ..........GTTCCGCGCCGAGCTCCTCGACACCGTGGGCACCGTCCGC 33665

   36 roGluProGlnProThrArgValIleLeuPheAsnLysProTyrAspVal 52
        ||:||  ::::|||
33666 CCGACCGCCGCCCG.............................. 33679

   53 LeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIl 69
                                               33679

   69 eProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluG 86
        ||||| :::|||
33680 ...........GCTGCCCTTCTTCCTCCACCGTGACCGCCGCG 33710

   86 lyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln 102
        ||||||
33711 GC.........................GCACGAGCCCGAAGG 33727

  103 ProGlyLysArgThrGlyLysTyrTyrValGlnValGluGly...... 117
        ||||||::::
33728 CCTGGACGGCCGCGTACTGGTACCGGAACATGCGCCAACCCGTGGAGTTCG 33777

  118 ...GluProAspAspAspAlaSerLeu........AlaLysLeuArgAsnG 130
        ::::||| |||||||
33778 CGTCCACCCTGCGGACGCGTGCTGCCGAGGGCCACCGCACCTTCGT..... 33823

  130 lyValThrLeuAsnAspAspArgProThrLeuProAlaGlyIleGlu.... 144
        :::|||||||||  |||||||
33824 .....CGAGATGGGCCCCACCCCTGCTGGCGCCGCGGATCGA 33862

  145 ..........ArgValAsnGluProGluTrpLeuTrpPr 154
        ||||::::::||
33863 CGAGGTCGCCGGCCGCGAGGGCGTGCACGCCACCGC............CC 33900

  154 oArgAsnProProIleArgGluArgLysSerIleProThrSerTrpLeuL 171
        |||||||::: |||::: |
33901 TCGCCACCCCTCCACCGGGCTCCGGGCGCCTGGACCGGTTCCGCTCCTC. 33949

  171 ysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAla 187
        |||    ::|||   |||
33950 ..........GGTGGGCGCCGCGTTCGCCCACGGAGTACGGGT 33982

  188 HisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly 202
        :::::|||  ||| ::|||
33983 CGACTGGGACGCCCT..CTTCGAGGGCTCCGACCCGCCGCCGGGT 34024
```

```
33679  ......................................................  33679

   69  eProValGlnGlyValTyrAlaGlyLyArgLeuAspArgAspSerGluG  86
             |||||    |||  ||||||  ::||
33680  ..............GCTGCCCTTCTCTCCACCGTGACCGCGCG  33710

   86  lyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln  102
       ||                                |||||    :::
33711  GC.....................GCACGAGCCCCGAAGG  33727

  103  ProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGly......  117
       |||||||::||::::    :::::||||||||:::||||
33728  CCTGGACGCCGCGTACTGGTACCGGGAACATGCGGGAACCCGTGGAGTTCG  33777

  118  ....GluProAspAspAlaSerLeu.........AlaLysLeuArgAsnG  130
          ::|||    |||||||:::                   ||||||
33778  CGTCCACCCTGCGGACCCTGCTGCGCGCAGGGCCACCCCACCTTCGT....  33823

  130  lyValThrLeuAsnAspGlyProThrLeuProAlaGlyLyIleGlu.....  144
          :::||||||||:::    ||||||||                 :::
33824  ...........CGAGATGGGGCCCGCACCCCCTGCTGGGCGCCGGATCGA  33862

  145  .........ArgValAsnGluProGluTrpLeuTrpPr  154
                   ||::::::::::          ||
33863  CGAGGTCGCCGAGGCCGAGGGCGTCGACGCCACCGC............CC  33900

  154  oArgAsnProProIleArgGluArgLysSerIleProThrSerTrpLeuL  171
       ||||::||||||    |||  ||:::                   |||
33901  TCGCCACCCTCCACCGCGGCTCCGGGCGGCCTGGACCGGTTCCGGCTCCTC.  33949

  171  ysIleThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAla  187
             |||    :::|||                   |||  |||||||
33950  ...............GGTGGGCGCCGGTCGCCCACGGAGTACGGGT  33982

  188  HisValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly  202
       ::::::|||  |||  |||||||    ::||||          |||
33983  CGACTGGGACGCCCCT...CTTCGAGGGCTCCGGCGCCCGGGGT  34024
```

OM of: US-09-252-691-7056 to: EST:*   out_format : pfs

Date: Jun 10, 2000 12:34 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09252691/runat_05062000_101649_585/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09252691
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: EST:*
Database sequences: 4857316
Database length: 2026611650
Search time (sec): 609.850000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len		Documentation
em_est20:AW017830	+	449.00	905.22	1.3e-41	295	!	AW017830 614064A09.x1 614 - roc
gb_est26:AI373988	+	392.00	789.70	3.4e-35	283	!	AI373988 SWOvAFCAP32B02SK Onchc
gb_est28:AI488035	+	129.50	249.02	4.5e-05	510	!	AI488035 EST246357 tomato ovary
gb_est37:AI993553	+	95.00	180.17	0.3063	452	!	AI993553 701496622 A. thaliana,
gb_est18:AA701638	-	92.00	167.09	1.64	844	!	AA701638 zi43e01.s1 Soares.feta
gb_est27:AI415775	+	89.00	169.11	1.87	408	!	AI415775 fb26h05.y1 zebrafish W
gb_est39:AW120050	+	87.00	160.40	3.87	618	!	AW120050 614002E03.y1 614 - roc
gb_est19:AA798012	+	84.00	151.47	12.16	796	!	AA798012 vy05b12.r1 Stratagene
gb_est31:AI681127	-	83.00	154.76	7.97	494	!	AI681127 tx44c01.x1 NCI_CGAP_Lu
gb_gss6:AQ871305	-	82.50	148.35	18.14	801	!	AQ871305 nbeb00042708f CUGI Rice
gb_est46:AV398064	+	82.00	147.95	19.10	758	!	AV398064 AV398064 Bombyx mori c
gb_est20:AA816992	+	82.00	147.27	20.82	805	!	AA816992 LD21944.5prime LD Dros
gb_gss15:AQ570573	+	81.50	151.31	12.41	512	!	AQ570573 HS_5362_B1_G11_T7A RPC
gb_est4:R65390	-	81.50	151.27	12.48	514	!	R65390 13894 Lambda-PRL2 Arabid
gb_est6:N58448	-	81.50	149.56	15.53	599	!	N58448 yv71e03.s1 Soares fetal
gb_est10:AA131370	-	81.50	149.47	15.70	551	!	AA131370 zo08c08.s1 Stratagene
gb_est38:AW066025	+	81.00	147.03	21.49	686	!	AW066025 687005A04.y1 687 - Ear
gb_est14:AA391680	-	81.00	144.87	28.33	832	!	AA391680 LD11613.5prime LD Dros
gb_est6:N55536	+	80.50	150.33	14.07	466	!	N55536 yv49a08.s1 Soares fetal
gb_est16:AA540892	+	80.50	149.97	14.73	481	!	AA540892 LD20751.5prime LD Dros
gb_est30:AI665224	-	80.50	147.56	20.07	597	!	AI665224 605008E01.x1 605 - Enc
gb_est31:AI667933	+	80.50	147.28	20.79	612	!	AI667933 TENG0888 T. Cruzi epim
gb_est31:AI677023	-	80.50	147.05	21.43	625	!	AI677023 605048C11.x1 605 - End
gb_est42:AW161632	+	80.50	146.07	24.28	682	!	AW161632 au68c06.x1 Schneider f
gb_est16:AA595396	-	80.00	148.37	18.09	507	!	AA595396 no36g07.s1 NCI_CGAP_Pr
gb_est30:AI658860	+	79.00	148.04	18.86	435	!	AI658860 tt98e11.x1 NCI_CGAP_Pr
gb_gss3:B51028	+	79.00	147.39	20.50	461	!	B51028 CIT9798SK-30G20.TV CIT978
gb_est14:AA435263	+	79.00	145.67	25.57	538	!	AA435263 vd49c11.s1 Knowles Sol
gb_est46:AW348346	+	79.00	144.66	29.11	589	!	AW348346 GM210002A13A3R Gm-r102
gb_est27:AI462356	-	79.00	142.05	40.68	744	!	AI462356 ub72b12.x1 Soares.mamm
gb_gss2:CNS014G6	-	79.00	138.32	65.64	1039	!	AL104112 Drosophila melanogast
gb_est47:AW436116	+	78.00	150.27	14.18	297	!	AW436116 75497 MARC 2PIG Sus sc
gb_est31:AI682440	+	78.00	146.61	22.65	412	!	AI682440 wc53d01.x1 NCI_CGAP_Pr
gb_est27:AI442518	-	78.00	145.63	25.70	450	!	AI442518 sa32e10.y1 Gm-c1004_Gl
gb_est45:AW245119	+	78.00	144.28	30.58	508	!	AW245119 2822678.3prime NIH_MGC
gb_est47:AT002659	+	78.00	142.80	36.97	580	!	AT002659 AT002659 POSLM01 Pleur
gb_est37:AW017796	-	78.00	142.68	37.52	586	!	AW017796 614063E12.x1 614 - roc
gb_est38:AW076286	+	78.00	142.49	38.44	596	!	AW076286 614063E12.y1 614 - roc
gb_gss14:AQ446999	+	78.00	142.34	39.18	604	!	AQ446999 mgxb0006218f CUGI Rice
gb_gss16:AW005576	+	78.00	142.07	40.58	619	!	AW005576 wz87h06.x1 NCI_CGAP_Br
gb_gss10:AQ161810	+	78.00	141.63	42.95	644	!	AQ161810 mgxb0009W06r CUGI Rice
gb_est19:AA246230	+	78.00	140.14	52.00	736	!	AA246230 LD06020:5prime LD Dros

gb_est22:AI068365	+	78.00	137.33	74.50	946	!	AI068365 mgae0002aA07f Magna
gb_est7:AW33234	+	77.50	144.71	28.92	446	!	AW33234 sh54b07.y1 Gm-c1015
gb_est47:AW433188	+	77.50	143.57	33.47	494	!	AW433188 sh53e07.y1 Gm-c1015

seq_name: em_est20:AW017830

seq_documentation_block:
ID AW017830 standard; RNA; EST; 295 BP.
XX
AC AW017830;
XX
SV AW017830.1
XX
DT 14-SEP-1999 (Rel. 61, Created)
DT 14-SEP-1999 (Rel. 61, Last updated, Version 1)
XX
DE 614064A09.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
DE mRNA sequence.
XX
KW EST.
XX
OS Zea mays
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Zea.
XX
RN [1]
RP 1-295
RA Walbot V.;
RT "Maize ESTs from various cDNA libraries sequenced at Stanford University";
RL Unpublished.
XX
DR MaizeDB; Probe/236176; p-std614064A09.
XX
CC Contact: Walbot V
CC Department of Biological Sciences
CC Stanford University
CC 855 California Ave, Palo Alto, CA 94304, USA
CC Tel: 650 723 2227
CC Fax: 650 725 8221
CC Email: walbot@stanford.edu
CC Plate: 614064 row: A column: 09.
XX
FH Key   Location/Qualifiers
FH
FT source   1..295
FT   /db_xref="taxon:4577"
FT   /db_xref="ESTLIB:1834"
FT   /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
FT   EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
FT   Lab (LM)"
FT   /organism="Zea mays"
FT   /cultivar="W23"
FT   /clone_lib="614 - root cDNA library from Walbot Lab"
FT   /tissue_type="root"
FT   /dev_stage="3-4 days old"
FT   /lab_host="XLOLR"
XX
SQ Sequence 295 BP; 80 A; 76 C; 76 G; 63 T; 0 other;

alignment_scores:
    Quality: 449.00   Length: 97
    Ratio: 4.726   Gaps: 0
    Percent Similarity: 97.938   Percent Identity: 88.660

alignment_block:
US-09-252-691-7056 x AW017830

Align seg 1/1 to: AW017830 from: 1 to: 295

US-09-252-691-7056 from: 1 to: 295

      21 HisArgValGluArgPheSerSerArgGlnAlaThrArgArgThrProGl 37
         |||:::::||||||||:::::|||||||||||:::||||||||||||||||

```
alignment_scores:
                      Quality:   392.00        Length:      80
                        Ratio:    5.227          Gaps:        0
   Percent Similarity:  93.750    Percent Identity: 91.250

alignment_block:
US-09-252-691-7056 x AI373988

Align seg 1/1   to: AI373988   from: 1  to: 283

123 SerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLe 139
    :::|||    ||||||||||||||||||||||||||||||||||||||||
  3 ACACTTGAAGCCTTGCGCAATGGCGTAACCTTAAATGATGGCCCTACCCT 52

139 uProAlaGlyIleGluArgValAsnGluProGluTrpProLeuTrpProArgA 156
    ||||||||   ||| ||||::||||  ||||||||||||  |||||||||||
 53 GCCCGCCGGCGCGGAGCTGGTTGACGAACCCGCCGGCTGGTTATGGCCGCGGA 102

156 snProProIleArgGluArgLysSerIleProThrSerTrpLeuLysIle 172
    ||||||||||||||||||||||||||||||||||||||||||||||||||
103 ATCCACCAATTCGTGAACGCAAAAGTATTCCCACCAGCTGGCTGAAGATC 152

173 ThrLeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisVa 189
    |||||||||||||||||||||||||||||||||||||||||||||||||||
153 ACCTTATATGAAGGACGTAATCGCCAGGTGCGCCATGACCGCCCATGT 202

189 lGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly 202
    ||||||||||||||||||||||||||||||||||||||||
203 TGGCTTCCCCACCCTGCGACTGATTCGCTATGCGATGGGT 242

seq_name: gb_est28:AI488035

seq_documentation_block:
LOCUS       AI488035       510 bp     mRNA           EST       29-JUN-1999
DEFINITION  EST246357 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED19M8, mRNA sequence.
ACCESSION   AI488035
VERSION     AI488035.1   GI:4383406
KEYWORDS    EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 510)
AUTHORS     Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
            Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
            Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE       Generation of ESTs from tomato carpel tissue
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137718.
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1..510
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED19M8"
                     /clone_lib="tomato ovary, TAMU"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /note="vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
```

```
  5 CACCAGGTTAAGCCATTCAGCTCGCAACGTCTACCAGGCGTAAACCTGA 54

 37 uProGlnProThrArgValIleLeuLeuPheAsnLysProTyrAspValLeuP 54
    | |||||||||||||||||||||||||||:::||||||||||||||||||||
 55 AAACCAGCCCACGCGTGTGATCCTGTTCAATAAACCCTACGATGTTCTTC 104

 54 roGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIlePro 70
    ||||||||||||||||||||||||||||:::||||||:||||||||
105 CGCAGTTCACCGATGAAGCCGGACGCAAAACATTAAAGAATTCATCCCG 154

 71 ValGlnGlyValTyrAlaAlaGlyTyrAlaGlyLeuAspArgAspSerGluGlyLe 87
    |||||||||||||||||||||||||||||||||||||||||||||
155 GTTCAGGGGTGTTATGCACGAGGTCGCCCTGACCGGATAGCGGAAGGGGT 204

 87 uLeuValLeuThrAsnAspGlyValLeuGlnValLnAlaArgLeuThrGlnProG 104
    ||||||||||||||||||||||||||||||||||||||||||||
205 GCTGGTGCTGACCAATAACGGCGTTGCAGGCGCGTTAACCCAGCCGG 254

104 lyLysArgThrGlyLysIleTyrTyrValGlnValGlnGly 117
    |||||||||||||||||||||||||||||||||||||
255 GTAAACGCACCGGAAAAAATCTATTATGTGCAGTGGAAGGT 295

seq_name: gb_est26:AI373988

seq_documentation_block:
LOCUS       AI373988       283 bp     mRNA           EST       15-JAN-1999
DEFINITION  SWOvAFCAP32B02SK Onchocerca volvulus adult female cDNA
            (SAW98MLW-OvAF) Onchocerca volvulus cDNA clone SWOvAFCAP32B02 5',
            mRNA sequence.
ACCESSION   AI373988
VERSION     AI373988.1   GI:4160024
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
ORGANISM    Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;
            Filarioidea; Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 283)
AUTHORS     Lizotte-Waniewski,M. and Williams,S.A.
TITLE       Genes expressed in adult female stage of Onchocerca volvulus
JOURNAL     Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797276.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1..283
                     /organism="Onchocerca volvulus"
                     /db_xref="taxon:6282"
                     /clone="SWOvAFCAP32B02"
                     /clone_lib="Onchocerca volvulus adult female cDNA
                     (SAW98MLW-OvAF)"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Filarial nematode parasite of humans. Two adult
                     female worms of Onchocerca volvulus were isolated from
                     consenting patients and quick frozen. Adult female mRNA
                     was converted to double-stranded cDNA using reverse
                     transcriptase and oligo(dT) followed by RNase H and DNA
                     pol I. The library has 7 x 10E5 independent recombinants
                     and the average  insert size is ~1100bp.The library was
                     constructed by Michelle Lizotte-Waniewski with worms
                     provided by Dr. Sara Lustigman. The library is available
                     from Dr.Steven A. Williams, email: genome@smith.edu."
BASE COUNT       63 a      86 c      74 g      60 t
ORIGIN
```

XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lamda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

```
BASE COUNT    151 a    111 c    113 g    135 t
ORIGIN

alignment_scores:
        Quality:    129.50        Length:    137
        Ratio:      1.660         Gaps:      4
Percent Similarity: 56.934   Percent Identity: 26.277

alignment_block:
US-09-252-691-7056 x AI488035

Align seg 1/1  to: AI488035   from: 1 to: 510

    34 ArgThrProGluProGlnProThrArgValIleLeu...PheAsnLysPr 49
       |||  ||:::   ||||||||||||||||  |||   ::|||||||
    91 CGCCTTCCTAAGAAATTGCCTACAAAGGTCTATCTTGCACTAAACAAGCC 140

    49 oTyrAspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuL 66
       |  :::   :::  ::::::::: ::  :||||:||:::::
   141 AAAAGGGTACATATGCTCATCTGGGGAGAAAGAAACTAAGTCAGTCATGT 190

    66 ys........AspPheIle.............................. 69
       ||||||||
   191 CCCTTTTGATGACTTTATAAAGAGTGGGATAAAAGGCATCCTGGACAA 240

    70 ProValGlnGlyValTyrArgAlaAlaGlyArgLeuAspArgAspSerGluGl 86
       |||      :::::::::::::::::::::||||||||
   241 CCGAAACCTCGGCTCTTTACAGTTGGCAGACTTGATGTTGCCACGACTGG 290

    86 yLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnP 103
       ||||||||||||||||||  :::             ::::||
   291 CTTGATTATTGTGACCAATGATGGGGAGTTCACGCGCCACCAGATTTCACATC 340

   103 roGlyLysArgThrGlyLysIleTyrTyrValGlnValGlyGluPro 119
       |||  :::::::::||||||||||||
   341 CTTCATCTAATTGTCAAAGGAATACATTGCAACTATCGACGGTGAAGTT 390

   120 AspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGl 136
       |||    :::|||   :::|||:||:::
   391 CATAAGCGACACTTGATAGCCATTAGTGAGGGAACAATATTGACGGTGT 440

   136 yProThrLeuProAlaGlyIleuGluArgVal...AsnGluProGluTrpL 152
       |||  ::::::|||  :::  :::::||||:|||
   441 CCATTGCACCCCAGATAAATGTTGAACTACTACCAGGGCCAGCCTGACTTAT 490

   152 euTrpProArg 155
       ||||||
   491 CAAGACCTCGT 501

name: gb_est37:AI993553
```

Tan,R., Rose,M., Warren,B., Ton.,B., Kastury,K., Borillo,C.,
Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S.,
Nobriga.,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and
Hanson,D.

```
FEATURES         Location/Qualifiers
     source      1..452
                 /organism="Arabidopsis thaliana"
                 /db_xref="taxon:3702"
                 /clone="701496622"
                 /clone_lib="A. thaliana, Ohio State clone set"
                 /note="cDNA library was made from selected clones from the
                 Arabidopsis thaliana Ohio State clone iset."
BASE COUNT    132 a    86 c    93 g    141 t
ORIGIN

alignment_scores:
        Quality:    95.00         Length:    111
        Ratio:      1.638         Gaps:      2
Percent Similarity: 52.252   Percent Identity: 26.126

alignment_block:
US-09-252-691-7056 x AI993553

Align seg 1/1  to: AI993553   from: 1 to: 452

    34 ArgThrProGluProGlnProThrArgVal...IleLeuPheAsnLysPr 49
       |||                ||| ||||||   :::|||||||||
   108 CGTATTCCTAAGAAACTTCCTCCAAAGTTTATTTTGCTCTGAACAAGCC 157

    49 oTyrAspValLeuProGlnPheThrAspGluAlaGlyArgSerThrLeuL 66
       |       :::     :::::   ::::||||:||:::::
   158 TAAAGGATAATATTGTCTTCCGGAGAGAAGAGATCAAATCTGCCATTA 207

    66 ysAspPhe.......................................Ile 69
       :::|||
   208 GTTTGTTTGACGAGTACTTGTCCTCTTGGGATAAAAGGAATCCGGGGACT 257

    70 ProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGl 86
       |||      ::::::::::::::::::||||||||
   258 CCCAAACCTCGCCTTTTACTGTTGGCGTCTTGATGTTGCCACACGGG 307

    86 yLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnP 103
       ||||||||||:||||||||||  :::             ::::||
   308 ATTGATAGTTGTTACAAATGATGAGATTTCGCCACAGAAACTTTCACATC 357

   103 roGlyLysArgThrGlyLysIleTyrTyrValGlnValGlyGluPro 119
       |||  :::::::::||||||||||
   358 CTTCATCTAGTTTACCGAAAGAATATTACCACGGTTGTCGGTGATATA 407

   120 AspAspAlaSerLeuAlaLysLeuArgAsnGly 130
       |||    :::|||   :::|||:||:::
   408 CACAAACGACACTTAATGGCTATCAGTGAAGGC 440
```

seq_name: gb_est18:AA701638

277 CCCATGGTATATATGGAGCCCTCACAAAAGCCCAGTGCCAGGGACTCC 228

119 OAspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspG 136
    |||||:|||:|||||||::::||||:::|||::   :|:
227 AGACTCAATTCTCCATTGGGAGCAGCAGAGAT......GTCCAGGTACAG 184

136 lyProThrLeuProAlaGlyLyIleGluArgValAsnGluProGluTrpLeu 152
    :|||| ||||||||||||:|||||
183 ATGCAAGTCTTGATGAGGAACTTGATCGAGTCAAGATGAGTTATTGGAGC 134

153 TrpProArgAsnProProIleArgGluArgLys................. 163
    ||| ||: |||||
133 TGGGCTTGGCCAGGAGTCTGGGACAAGGAAGCAGATTTTCCTGATTCT 84

164 ....SerIleProThrSerTrp 169
             |||:|||:|||
 83 GGCTCTAGCTTCCCTGCCAAGATTTGG 56

seq_name: gb_est27:AI415775

seq_documentation_block:
LOCUS   AI415775   408 bp   mRNA   EST   09-FEB-1999
DEFINITION fb26h05.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5', mRNA
            sequence.
ACCESSION   AI415775
VERSION   AI415775.1 GI:4259279
KEYWORDS   EST.
SOURCE   zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1 (bases 1 to 408)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE   WashU Zebrafish EST Project 1998
JOURNAL   Unpublished (1998)
COMMENT   On May 7, 1998 this sequence version replaced gi:3118831.
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 395.
FEATURES   Location/Qualifiers
     source   1..408
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF'"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCGCCCCTTTTTTTTTTTTTT3'];
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed

--- (second column) ---

KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 844)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE   WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1392833.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 113.
FEATURES   Location/Qualifiers
     source   1..844
                     /organism="Homo sapiens"
                     /db_xref="GDB:1333329"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:433560"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library. 1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT   205 a 214 c 213 g 212 t
ORIGIN

alignment_scores:
   Quality: 92.00   Length: 126
   Ratio: 1.438   Gaps: 3
Percent Similarity: 50.794   Percent Identity: 26.190
alignment_block:
US-09-252-691-7056 x AA701638/rev

Align seq 1/1 to reverse of: AA701638 from: 1 to: 844

 53 LeuProGlnPheThrAspGluAlaGlyArgSerThrLeuLysAspPheIle 69
    ||| ||| :::|||:|| ||| ::
418 CTGACAACATAATGGAGATGCAGCAGCTTCACTGACT........GT 378

 69 eProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluG 86
    :||| ||| ||| ::: ||| :::||| ||||||||||:::
377 GCCCGAGCAGTATGCCAGGCGCTTTCCCAAATTGGCCAGGGACTCCAACA 328

 86 lyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln 102
    :::||||| ||| ||| ||||| :::||| ||||||:::|:||
327 TTTTCTTACTGCCCTCCAACTCTGGCCATGTCACCAGCTTCGTGTCTCAG 278

103 ProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyGluPr 119
    ||| ||| ||| ::: :: ||| ::: ||| ||| ||| ||

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

```
BASE COUNT     130 a     100 c     109 g      69 t
ORIGIN

alignment_scores:
     Quality:      89.00              Length:        57
     Ratio:        2.405              Gaps:           1
Percent Similarity:  64.912     Percent Identity:  31.579

alignment_block:
US-09-252-691-7056 x AI415775

Align seg 1/1  to: AI415775  from:  1  to: 408

 120 AspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGl 136
     ::|||||||| ||     ::  ||||||||||    :||||:|:::||
   4 GAGGACGGCGATACTGAGAGCGGGAACGAGTTCGCGATAGTGAAGAAACGG  53

 136 yProThrLeuProAlaGlyIleGluArg........................ 145
     |||||:|||  ||| |||||||
  54 ACCAAGCGCCCGCAGGGAGCAGAGAGGAGGAGCACACGGCTGGGACCGCGA 103

 146 ..ValAsnGluProGluTrpProLeuTrpProArgAsnProProIleArgGlu 161
     ::|||||||:::::|||||||  :||  :||||:||||:::
 104 CGATAGACTCGATCGATTGGGGTCTGGCCGGCCGAGGAACCCCATGAGAGAC 153

 162 ArgLysSerIleProThrSer 168
     ::||||||||||||||||
 154 TCTCAGCAAATACCAACGAGC 174

seq_name: gb_est39:AW120050

seq_documentation_block:
LOCUS       AW120050      618 bp      mRNA         EST         22-OCT-1999
DEFINITION  614082E03.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW120050
VERSION     AW120050.1   GI:6095383
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1  (bases 1 to 618)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188325.
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614082 row: E column: 03.
FEATURES             Location/Qualifiers
     source          1..618
                     /organism="Zea mays"
```

```
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT     180 a     128 c     181 g     129 t
ORIGIN

alignment_scores:
     Quality:      87.00              Length:       193
     Ratio:        0.978              Gaps:          10
Percent Similarity:  46.114     Percent Identity:  21.762

alignment_block:
US-09-252-691-7056 x AW120050

Align seg 1/1  to: AW120050  from:  1  to: 618

  20 LysHisArgValGluArgPheSerSerArgGlnAlaThrArgArgThrPr  36
     ::||||||||||||     ::  |||    : :||||   ::
  84 AGAGAGAGTCGAATCAAGGCCGACCGTGGGCGGTCCAAGACCTACCC     133

  36 oGluProGlnProThr.........................ArgValIleLeuPheAsnL  48
     |||                                      :|||
 134 GCAGCAGGCTGGCACCGTCCGTAAGACAGCGGCTTCATCGTCATCAAGAACC    183

  48 ysProTyrAspValLeuProGlnPheThrAspGluAlaGlyArg.......  62
     :||  |||:||  ||||||||||:|||||:|:|:::||
 184 GTCCCTGCAAGGTTGTGGAGGTTTCTACCTCCAAGACTGGTAAGCCATGGC    233

  63 .................................................SerThrLe  65
                                                      ||
 234 CACGCCAAATGCCACTTTGTCGCCATAGACATATTCAATGGGAAAAAGCT    283

  65 uLysAspPheIleProValGln.......GlyValTyrAlaAlaGlyA  79
     |:||:||::|:|::|||                 ::  ||||||
 284 TGAAGAATTGTTCCTTCATCACACAACTGTGACATTCCGCATGTGAACC    333

  79 rg.......................LeuAspArgAspSerGluGlyLeuLeu...Val  89
     ||                       :::||        |||:::|||  ::
 334 GTACTGAGTATCAGCTGATTGATATTTCTGAGGATGGATTTGTGAGCCTT    383

  90 LeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnPro........ 103
     |||||||||:|::||              ||| :::
 384 CTTACTTCAGATGGCAACACTAGGATGATCTTAGACTCCCACTGATGA    433

 104 ...............................GlyLysArgThrGlyLysIleTyrT 112
                                    ||||||   |||:||:||
 434 GACTCTTGTGGCCCAGATTAAGGAAGGGTTCGAAAGCGGGAAGGGATCTTG    483

 112 yrValGlnValGlyGluProAspAspAlaSerLeuAlaLysLeuArg     128
     ||| |||:||::||     :::    |||  ||| :|||
 484 TTGTGACTGTCCAGTCCGCATGGGGAGGAGGCCAGATCTGCCGCCTGAAG    533

 129 AsnGlyValThrLeuAsnAspGlyProThr...LeuProAlaGlyIleGl  144
     ::|||                            |||||
 534 ............GATGTTGCCCCAAGTAACTTCCT.............    557

 144 uArgValAsnGluProGluTrpProLeuTrpProArgAsnProProIleArgG  161
     ||||||||||                      ||||||:::|||
 558 ..........TGGCTTTGGAGACGGAACCCTACTCTATCTC    588

 161 luArgLysSerIleProThrSerTrpLeu 170
     :: |||||||        :::|||:|||
 589 AACTTAAATCG...GAAAAGAAGTGGATC  614

seq_name: gb_est19:AA798012
```

```
seq_documentation_block:
LOCUS       AA798012      796 bp      mRNA      EST      10-FEB-1998
DEFINITION  vy05b12.r1 Stratagene mouse macrophage (#937306) Mus musculus cDNA
            clone IMAGE:1294559 5' similar to gb:S85655 PROHIBITIN (HUMAN);
            gb:X78682 M.musculus mRNA for B-cell receptor associated protein
            (MOUSE).; mRNA sequence.
ACCESSION   AA798012
VERSION     AA798012.1  GI:2860967
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 796)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152124.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mousees@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:675607
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1..796
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1294559"
                     /clone_lib="Stratagene mouse macrophage (#937306)"
                     /tissue_type="macrophage"
                     /dev_stage="WEHI-3 cell line"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. WEHI-3 cell line. Average insert size: 1.5 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT  193 a  208 c  229 g  165 t  1 others
ORIGIN

Align seq 1/1  to: AA798012  from: 1  to: 796

alignment_scores:
    Quality:      84.00          Length:         216
    Ratio:        0.832          Gaps:            11
Percent Similarity:  46.759   Percent Identity:  26.389

alignment_block:
US-09-252-691-7056 x AA798012

    9 ProGluAsnThrMetThrLysThrSerPheArgLysHisArgValGluAr   25
      ||| .::||||   |||:|||       |||       |||
   80 CCGCTCTCGACCACGAACGTACCAGT.....CATCACTGGTAGCAAAG    123

   25 gPheSerSerArgGlnAlaThrArgArgThrProGluProGln........   39
      |:::::::|||||       :::    |||       |||
  124 ACTTGCAGAACGTCAATATCACACTGCGAATCCTCTTCCGGCCGGTGGCA   173

   40 ......ProThrArgValIleLeuPheAsnLysProTyrAsp.......   51
            |||               |||||||
  174 GCCAGCTTCCTCGTATCTCACACAGCATTGGCAGGACTATGATGAGCGG   223

   52 ValLeuProGlnPheThrAspGluArgSerThrLeuLysAspPh         68
      |||||||||:::::||| ||| .:||:||||::::::
  224 GTCCTGCCTTCATCACCAGAGATCCTCAAGTCGGTGGTGGCTCGATT     273

   68 eIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerG   85
                                  |||||:||||
  274 C............................GATGCTGGAGAATTG.......   289

   85 luGlyLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThr  101
                                  .::||||     |||  :||:
  290 ..........ATTACCCAGCGAGACGTACAACCAGGCAGGTGACGCG     325

  102 Gln.ProGlyLysArgThrGlyLysIleTyrTyrValGlnValGluGlyG  118
      .::  |||    .::  .:||:::||||:
  326 ATGACCTCACAGAGCGAGCGGCAACATTGG................GCTT  360

  118 luPro.AspAspAlaSerLeuAlaLysLeuArgAsnGlyValThrLeuAs  134
        ||       ||||||:::::
  361 ATCCTGGATGACGTGTCCCTGCACACATCGACCTTCGGAAGGAGTTCAC  410

  134 nAspGlyProThrLeuProAlaGlyIleIleGluArgVal...AsnGluProG  150
                       :::::|  ||:|||||||||
  411 AGAGGCAGTAGAAGCAAACAGGTGGATCAGCAGGAGCCGAACGAGAGCCAG  460

  150 luTrpLeuTrpProArgAsnProProIleArgGluArgLysSer......  164
         .::   ||| |||:|:::||
  461 ATTTGTGGTGGAAAAGGCTGAGCAGCAGAAGAGGCAGCCATCATCCTCTG  510

  165 ..............IleProThrSerTrpLeuLy                  171
                       ||      :
  511 CTGAGGGTGACTCACAGGCCGATGAGCTGATCGCCCAACTCGCTGGCACG  560

  171 sIleThrLeuTyrGluGlyGlyArgAsnArgGlnValArgArgMetThrAlaH  188
         |||       ||||||| ||||:::::
  561 GCCGGACGACGACTGAATCGAGACTGCGGAACTTGGAGGGCACGACGGGG.  609

  188 isValGlyPheProThrLeuArgLeuIleArgTyrAlaMetGly         202
         ||||||| |||:::|||
  610 ..........ACATTGAGATCCAGCTCTCCCGCTCTCCGGA            639

seq_name: gb_est31:AI681127

seq_documentation_block:
LOCUS       AI681127      494 bp      mRNA      EST      16-DEC-1999
DEFINITION  tx4c01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272416 3'
            similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION   AI681127
VERSION     AI681127.1  GI:4891309
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1134367.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 760     Std Error: 0.00
```

```
              Seq primer: -40UP from Gibco
              High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source      1..494
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:2272416"
                 /clone_lib="NCI_CGAP_Lu24"
                 /tissue_type="carcinoid"
                 /lab_host="DH10B"
                 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                 modified polylinker; Plasmid DNA from the normalized
                 library NCI_CGAP_Lu5 was prepared, and ss circles were
                 made in vitro. Following HAP purification, this DNA was
                 used as tracer in a subtractive hybridization reaction.
                 The driver was PCR-amplified cDNAs from a pool of 5,000
                 clones made from the same library (cloneIDs
                 1414920-1417991 and 1520904-1522439). Subtraction by Bento
                 Soares and M. Fatima Bonaldo."       1 others

BASE COUNT      83 a     114 c     159 g     137 t     1 others
ORIGIN

alignment_scores:
           Quality:      83.00        Length:       107
           Ratio:        1.596        Gaps:         5
    ercent Similarity:   48.598       Percent Identity:  28.972

alignment_block:
US-09-252-691-7056 x AI681127  ..
Align seg 1/1  to: AI681127     from: 1  to: 494

  135 AspGlyProThrLeuPro............................AlaGlyIle 143
      |||||||||||||||||                             |||::::|||
  122 GACGGCCCACCCTTCCAGGGACACAGGGGACTTGGTTTCTGCTTCTCT 171

  143 eGluArgValAsnGluProGluTrpLeuTrpPro............ 154
      |||  ::||||   |||||| |||||
  172 AGAGGTACATGGGGACAAGCGTGGCTTCTGCCAGACTTGTTTAGCTGTG 221

  155 ...ArgAsnProIleArgGluArgLysSerIlePro.......... 166
      :::||   :||||::::::   |||                   ||
  222 ACCTGCAGGGCAGGCCCCTGGGGCTGCCACTTGGTTCACCAGATGGAAGG 271

  167 ThrSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValAr 183
      :::||   ::||::    :::||   |||::       |||||
  272 TCAGCATGGCTGNGCCACGCTTGTCTGAGAGGGTCCTGGCAGTGGGTACG 321

  183 gArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrA 200
      |||||| :::::::::::||||:|||:||||||
  322 CAGGATGTATGGCAGGGCCCTTTTTCCAGTGCTT...............G 356

  200 laMetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGlu........ 213
      ||:::||||   ::: ||||  :||||        ||||||
  357 CTCTAGGCCAGTCCTCGCAGGGACCCCTGCTCTGGGGAGAGAGGCAGGGT 406

  214 ...........TrpArgAsp 216
             ||| |||
  407 GTGCCCAGGCTTGGTCCGAT 427

seq_name: gb_gss6:AQ871305

seq_documentation_block:
LOCUS       AQ871305     801 bp    DNA               GSS        03-NOV-1999
DEFINITION  nbeb0042J08f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0042J08f, genomic survey sequence.
ACCESSION   AQ871305
VERSION     AQ871305.1  GI:6221756
KEYWORDS    GSS.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 801)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     On Feb 19, 1999 this sequence version replaced gi:4130187.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 17
            High quality sequence stop: 384.
FEATURES             Location/Qualifiers
     source      1..801
                 /organism="Oryza sativa"
                 /strain="Japonica"
                 /cultivar="Nipponbare"
                 /db_xref="taxon:4530"
                 /clone="nbeb0042J08f"
                 /clone_lib="CUGI Rice BAC Library (EcoRI)"
                 /tissue_type="Leaf"
                 /lab_host="E. coli DH10B"
                 /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                 Rice is the most important food crop in the world. Half of
                 the world population, especially those inhabiting highly
                 populated areas of the humid tropics and subtropics, rely
                 on rice as their primary source of carbohydrate.
                 Monocotyledonous rice is a diploid plant (2n=24) with a
                 haploid genome equivalent of 431 Mbp (Arumuganathan and
                 Earle, 1991). The relatively small genome of rice, three
                 times larger than that of Arabidopsis, makes it suitable
                 for genomic studies. In order to facilitate positional
                 cloning, physical mapping and genome sequencing of rice,
                 we have constructed a BAC library from Oryza sativa,
                 Nipponbare variety using EcoRI as the cloning enzyme. The
                 library contains 55,296 clones with an average insert size
                 of 121 Kb providing approximatley 15 haploid genome
                 equivalents. The deep coverage allows the isolation a
                 particular sequence with a probability of 99.9 %. Three
                 high density filters, each containing 18,432 clones
                 (doubly spotted), represent the whole library for colony
                 screening and can be requested from the Clemson University
                 BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT     167 a     248 c     236 g     148 t     2 others
ORIGIN

alignment_scores:
           Quality:      82.50        Length:       239
           Ratio:        0.868        Gaps:         12
    Percent Similarity:  39.749       Percent Identity:  22.176

alignment_block:
US-09-252-691-7056 x AQ871305/rev  ..
Align seg 1/1  to reverse of: AQ871305     from: 1  to: 801

   29 ArgGlnAlaThrArgThrProGluProGlnPro................ 40
      |||::||:|||||:|||||||||||
  729 CGCAGCTCGACTCGCCGGACAGGTGAGCCCCTACTCATCGATGCAGATT 680

   41 .......................ThrArgValIleLeuPheAsnLysPro. 49
                           |||||||  ::::::|||:|:|:|||
  679 CTCCCTCTTCCGGTCCCGTTGAACGAGATTTTGTTCTTTTCACGTCCGC 630

   50 .....................................TyrAspVal 52
                                           |||::::
```

Email: kmita@nirs.go.jp
method:uni-directional, sequence direction:sequenced from T3 primer
(5' -> 3')
Project='silkworm Genome Program in MAFF, and Research for the
Future Program in JSPS'. see 'SilkBase',
<http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.

```
FEATURES             Location/Qualifiers
     source          1..758
                     /organism="Bombyx mori"
                     /db_xref="taxon:7091"
                     /clone="NV021042"
                     /clone_lib="Bombyx mori ovary BmNPV infected; 2 hr after
                     inoculation"
                     /tissue_type="ovary"
                     /cell_type="BmN cultured cell"
                     /dev_stage="BmNPV infected; 2 hr after inoculation"
BASE COUNT      179 a     218 c     249 g     112 t
ORIGIN
```

alignment_scores:
```
        Quality:     82.00           Length:      254
        Ratio:        0.732          Gaps:         15
        Percent Similarity: 44.094   Percent Identity: 23.228
```

alignment_block:
US-09-252-691-7056 x AV398064

Align seg 1/1 to: AV398064 from: 1 to: 758

```
 24 GluArgPheSerArg.........................GlnAlaThrArgArgThrPr  36
    |||||                                   |||  ::|:::||
 30 GAGCGCGGGTGCCAGCCGCCTGCGTCGCCCCCGGCTCAACACCGACC                79

 36 oGluProGlnProThrArgValIleLeuPheAsnLysProTyrAspValL             53
    |::||| |:|:|| |||:::|                            |
 80 GCCGCCAGAGCCGCCCTCGAGCC...........................            101

 53 euProGlnPheThrAsp.............................GluAla          60
    :::|                                          ||:::|
102 ..CCGCAGCCCAGTCCCGAAACCGAACCGGAGGTGAACGGCGGCGGAGGGC          149

 61 GlyArgSerThrLeuLysAspPhe.........IleProValGlnLeuGl            73
    |||||||:::|||:::|||                  ::|||:::||
150 GGGAAGGCGACGCTGCCGTGTATATTCCACGGGAACATGCCGCCCAAGAT          199

 73 yValTyrAlaAlaGlyArgLeuAspArgAsp...SerGluGlyLeuLeuV            89
    :::||:                          |||  |||||||||
200 CACTTGGAGGAAAGGAGAAATCACTATCGACGGGCAGCGAGGCCGGACGC          249

 89 aILeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLys          105
    :::||:                                      :::|||
250 GCGTGCTGTCGGACGGCACGCTGGAG......ATCGTGTCGCTGTACCGG          293

106 ArgThrGlyLysIleTyrTyrValGlnValGluGlyGluProAspAspAl          122
    :::|||                                :::||| |||:||
294 AACGACACCGGCGTCTACATCTGCATCGGCTGAGAACGAG........TT          334

122 aSerLeuAlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrL          139
    :::|||                          :: :::|||||
335 CGGCATCAGCCAGCAGGAGATCCACCTGCAAGTGAACGATCCGGTGAGGA          384

139 euProAlaGlyIleGlu...ArgValAsnGlu.................            148
    :::|||||:::|||     :::|||
385 CGCCGGTCGGCATCGGCGGAGGAGACAAAACGRAGTAATAACGGGGAGATG         434

149 ................................................ProGl         150
                                                    |||||
435 GGCAGACCGCTCGTCTTGCGCTGCCTCGTGTACGGATACCCGACCCCGGA          484

150 uTrpLeuTrpProArg.....AsnProProIleArgGluArgLysSerI           165
    :::||| |||  |||      ||| |||::
485 AATCTTCTGGTACAGAGGACTTAACGGTCCCATG..............G           519
```

```
629 TTCGGATTTAGTGTGTTTATGCCATTGCCCCGGGGCGTCTACGAGTGT           580

 53 LeuProGlnPheThrAspGluAlaGly..............                    61
    ||:::||
579 GCGTGCAGGTCCGATGAGGAAAGCGGCGGCGAAATCGACTGGACAACCT          530

 62 ...........................................ArgSerThrL        65
                                               |||  |||
529 GGGGTTCGGGCTGACGCCGACCGACTACATGTACGTCATGCGGTGCTCGC          480

 65 euLysAspPhePheProValGlnGlyValValTyrAlaAlaGlyArgLeuAsp        81
    ||:::|||  ||:::||:::||:
479 TGGAGGAC..........GGCGTCTTCTCCGGCGGAGCTCAGC                 445

 82 Arg...........................AspSerGluGlyLeuLeu........     88
    |||                           ||:::||
444 CGCTACGGCAACATCGAGCTCAGCCCCTCCCGGGGTCATCAACTACGG           395

 89 ........VaILeuThrAsnAspGlyValLeuGlnAlaArgLeuT              101
           ||||||
394 CCAGGTCGTCACGTTCTCCAGTCACACACATGTCCATGCAACGAAACG           345

101 hrGlnProGlyLysArgThrGlyLys................                  109
    :::|||||  ||||||:::|||
344 TCGATCCAGGCGACCGACTAACAAGTGAACGAGGGTGATGGCTGCGTGC          295

110 ......IleTyrTyrValGlnValGluGlyGluProAspAspAlaSerLe          124
          :::|||
294 GTACGTACGTATGTGCTC.........GATGATCAGGGGCT                   263

124 u.AlaLysLeu......ArgAsnGlyValThrLeuAsnAspGlyProThr          138
     |:|||         |||  :::||| ||| |||:::
262 CTTCGAGGGTCTGAGGCGTACAGGCGCGGCGAACCAACAGGGGTCTACA          213

139 LeuProAlaGlyIleGluArgVal...............AsnGluProGl          150
    :::|||||                               |||
212 TGCTGTTCCGGCGGCGGAGGAGAACGCGCCGCGGATCGCACGGCGCCGAG         163

150 uTrpLeu.................TrpProArgAsnProProIleArgG            161
    :::|||                 ||| :::||| |||
162 CGCATGTGCATGCCGTCGCGTGGTCGGTGGAGCAGTTCGTCCACGCGTCAA        113

161 luArgLysSerIleProThrSerTrpLeuLysIleThrLeuTyrGluGly          177
    :::|||:::|||                      :::||| |||
112 GCAGACCGTCCTCGCCAACCGCCGCTGGGTACGTAAGGCGGTACGACGAA          63

178 ArgAsnArgGlnVal                                            182
    |||||
 62 CGTAACACGACCGTC                                            48
```

seq_name: gb_est46:AV398064

Align seg 1/1   to: AA816992   from: 1 to: 805

```
 46  PheAsnLysProTyrAspValLeuProGlnPheThrAspGluAlaGlyAr  62
     |||||||||                      |||               |||
 57  TTCAATAAA...TATCAA..............ACGACGACGAGTGCCCG  88

 62  gSerThrLeuLysAspPheIleProValGlnGlyValTyrAlaAlaGlyA  79
     |:||  |||::  ||||:                |||          ||||
 89  CAATGAAATTCCGATCTGCACTTTCTGCGGGGTCCTCACGTGGCC  138

 79  rgLeuAspArgAsp.........SerGluGlyLeuLeuValLeuThr  91
     |::||                         ||| ||:|||||||:|| |||
139  GCTTCCGCGAGGCGAGGTCAAAGTTGAGGAGGGTGTTCTGGTGGCGACG  188

 92  AsnAspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGl  108
     ||:|                         |||
189  GTGGACAACTTCAAGCAGCTGATTGCCACAACGAG.............  224

108  yLysIleTyrTyrValGlnValGlu.GlyGluProAspAspAlaSerLeu  124
     ::|||  ||||||| ::|  |||:|::                |||||::
225  ..........TTCGTGCTGGTCGAGTTCTACGCCCATGGTGCGGACACT  264

125  AlaLysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAl  141
     |||||||||                    ::|    |||||||::||||::
265  GCAAAGCTC..............TGGCTCCCGAGTACGCCAAGGCTCCCAG  302

141  aGlyIleGluArgValAsnGluProGluTrpProLeuTrpProArgAsn...P  157
     ::|   |||   |||||                 ||||||||||    |||
303  CAACTGGCCGAGAAGGAGTCGCCCATCA...AGCTGGCCAAGGTCGATGC  349

157  roProIleArgGluArgGlyLysSerIleProThrSerTrpLeuLysIleThr  173
     ||||| |||                      ::|||||||     ::  |||
350  CACCGTCGAGG..................GTGAGCTGGCCGAGCAGTACG  381

174  LeuTyrGluGlyArgAsnArgGlnValArgArgMetThrAlaHisValGl  190
     |||                                                ||
382  CC.............................GTCCGT.............GG  391

190  yPheProThrLeuArgLeuIleArgTyrAla..............MetG  202
     |:|||||||||||||||:::||||
392  CTACCCCACTCTGAAGTTCTTCCGCAGCGGCTTCCAGTGGAGTACAGCG  441

202  lySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrpArgAsp  216
     ||:::  |||||||||                 |||   ::|||:||
442  GTGGCCCCACGCTGGATCATCATTGCCTGGGTGACCAAGAAGAC  485
```

seq_name: gb_gss15:AQ570573

```
seq_documentation_block:
LOCUS       AQ570573     512 bp    DNA         GSS       01-JUN-1999
DEFINITION  HS_5362_B1_G11_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=938 Col=21 Row=N, genomic survey sequence.
ACCESSION   AQ570573
VERSION     AQ570573.1  GI:4963793
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     On Mar 23, 1999 this sequence version replaced gi:3325394.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
```

```
165  leProThrSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGln  181
     ::|||  |||                  |||||||||||::||:::::|
520  TGCCATACAGC........AGTACATTGTACGAAGCCAGAGAAAACGTT  560

182  ValArgArgMetThrAlaHisValGlyPheProThrLeuArgLeuIleAr  198
     ||||                                      ::|||||||
561  CTTCTA..................ATAAGACAACTGATCGA  583

198  gTyrAlaMetGlySerTyrThrLeuAspSerLeuAlaAsnGlyGluTrp.  214
     ::||:::|||||| | |:::::  ||
584  CGAGGCGCTCGGGAATACGGCGTGTCAAGCATAC...AACGGAGGAA  630

215  .................................ArgAspValThr  218
                                       ::|||  |||
631  GCCCGGCCACGTTACTGATGGAGGTTCGGGCGTACACAAACAAGACGATACG  680

219  ProLysGluAsn  222
     :::|||||||:|||
681  CCCTCTGACAAC  692
```

seq_name: gb_est20:AA816992

```
seq_documentation_block:
LOCUS       AA816992     805 bp    mRNA        EST       25-NOV-1998
DEFINITION  LD21944 5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD21944 5prime similar to U18973: Pdi
            FBgn0014002 PID:g622993 SWISS-PROT:P54399, mRNA sequence.
ACCESSION   AA816992
VERSION     AA816992.1  GI:2886601
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 805)
AUTHORS     Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE       BDGP/HHMI Drosophila EST Project
JOURNAL     Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2044731.
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 219 row: D column: 8
            High quality sequence stop: 518.
FEATURES             Location/Qualifiers
     source          1..805
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD21944"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2."
BASE COUNT   183 a    219 c    236 g    167 t
ORIGIN
```

alignment_scores:
```
Quality:          82.00          Length:   182
Ratio:            0.953          Gaps:      12
Percent Similarity: 47.253       Percent Identity: 29.670
```

alignment_block:
US-09-252-691-7056 x AA816992

```
                    Tel: (206) 616-3618
                    Fax: (206) 616-3887
                    Email: jwallace@u.washington.edu
                    Clones are derived from the human BAC library RPCI-11. For BAC
                    library availability, please contact Pieter de Jong
                    (pieter@dejong.med.buffalo.edu). Clones may be purchased from
                    BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
                    or from Resear h Genetics (info@resgen.com). BAC end Web Server:
                    http://www.htsc.washington.edu
                    Plate: 938  row: N  column: 21
                    Seq primer: T7
                    Class: BAC ends
                    High quality sequence stop: 512.
source              1..512
                    Location/Qualifiers
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Plate-938 Col-21 Row-N"
                    /clone_lib="RPCI-11 Human Male BAC Library"
                    /sex="male"
                    /note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT     151 a     158 c     108 g      88 t      7 others
ORIGIN

alignment_scores:
        Quality:      81.50         Length:      103
        Ratio:         1.538        Gaps:          7
Percent Similarity:   51.456    Percent Identity:  33.981

alignment_block:
US-09-252-691-7056 x AQ570573

Align seg 1/1  to: AQ570573     from: 1  to: 512

107 ThrGlyLysIleTyrTyrValGlnValGluGlyProAspAspAlaSe 123
    |||::||||||| ||||||   :::|||||::: ::|||
 99 ACAAGCAGCATCCTTACGTTTTA........CAACCAGAAGCCGGGTC 139

123 rLeuAlaLysLeuArgAsnGly...................... 130
    | |||| |||:::||| |||||
140 GTCTGTAAAAACCACGACAGGGTGAGAAACGGACACAATATCCATGGCTCG 189

131 ..ValThrLeuAsn......AspGlyProThrLeuProAlaGlyIleGlu 144
    ::|||:||||||  :::::|  ||||||| |||||||   ||||||
190 CGGCCACTATAAACCGGCCCAGCCACCGACGCGCTTCCA......ATTGAG 233

145 ArgValAsnGluProGluTrpLeu.....TrpProArgAsnProProIl 159
    |||  |:::|||||||    :::||||||| |||||||
234 AGACCGCAGAGCCCCAGAGACTGCTACAAGATGGCCTCCGTACCCGCCATA 283

159 eArg........GluArgLysSerIleProThrSerTrpLeuLysIleT 173
    ||| ::|||||  ||| ||| ||| |
284 AAGACTCTATTAGGACACACCCCAGCAAACCTCGCTTATGGGGAAAGCCA 333

173 hrLeuTyrGluGlyArgAsn...ArgGlnValArgArgMetThrAlaHis 188
    ||  ::: |||  :::::|||   |||||||   ||||||
334 CAGCCGGACGCAGAGCTAACGCCTCGAAAGCCAGAGCTAATGGAGCGCAC 383

189 ValGlyPhe 191
    |||
384 CTCACTTTT 392

seq_name: gb_est4:R65390

seq_documentation_block:
LOCUS     R65390     514 bp    mRNA         EST        09-JAN-1998
DEFINITION  13894 Lambda-PRL2 Arabidopsis thaliana cDNA clone 170K5T7, mRNA
            sequence.
ACCESSION   R65390
VERSION     R65390.1   GI:937860
KEYWORDS    EST.
SOURCE      thale cress.
```

```
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1 (bases 1 to 514)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
            McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
            Retzel,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Apr 14, 1993 this sequence version replaced gi:838028.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES    Location/Qualifiers
  source    1..514
            /organism="Arabidopsis thaliana"
            /strain="var columbia"
            /db_xref="taxon:3702"
            /clone="170K5T7"
            /clone_lib="Lambda-PRL2"
            /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
            Lambda PRL2 is a cDNA library derived from equal
            quantities of 4 pools of mRNA. The mRNA sources were 1) 7
            day germinated etiolated seedlings; 2) tissue culture
            grown roots; 3) staged plants half with 24 hour light
            cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
            same plants as 3 but aerial tissue (stems, flowers and
            siliques. The vector is BRL's lambda Zip-Lox. The cDNA
            inserts were directionally cloned with Sal-Not arms using
            oligo dT primed cDNA."
BASE COUNT    137 a     101 c     104 g     152 t     20 others
ORIGIN

alignment_scores:
        Quality:      81.50         Length:       95
        Ratio:         1.567        Gaps:          4
Percent Similarity:   54.737   Percent Identity:  29.474

alignment_block:
US-09-252-691-7056 x R65390

Align seg 1/1  to: R65390   from: 1   to: 514

 34 ArgThrProGluProGlnProThrArgVal...IleLeuPheAsnLysPr 49
    ||| ||| ||| ::: ||| |   ::::|||| ::|||||||
108 CGTATTCCTAAGAAACTTCCTCCAAAGGTTTATTTTGCTCTGAACAAGCC 157

 49 oTyrAspValLeuProGlnPheThrAspGluAlaGlyAlaGlyTyrSerThrLeuL 66
    |      :::  :::|||  ::::|||:::|||||
158 TAAAGGATATATTTGTCCTTCCGGAGAGAAAGAGATCAAATCTGCTATTA 207

 66 ysAspPhe..................................Ile 69
    :: |||
208 GTTTGTTTGACGAGTACTTGTCCCTTGGGATAAAGGAATCCGGGGACT 257

 70 ProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSerGluGl 86
    |||        :::|||||||||||||||||
258 CCCAAACCTCGCCTTTTACTGTTGGTCGTCTTGATGTTGCCACAACGGG 307

 86 yLeuLeuValLeuThrAsnAspGlyValLeuGlnAlaArgLeuThrGln. 102
    ||||:||:||:|||||||||||||   :::|||||
308 ATTGATAGTTGTTACAAATGATGGAGATTTCGCACAGAAACTTTCACATC 357
```

```
103 ProGlyLysArg...ThrGlyLysIleTyrTyr 112
    ||:|:::|||   |||   ||||||||
358 CCTTCATCTAGGTTTACCGNAAGNATATTAC 390
```

seq_name: gb_est6:N58448

seq_documentation_block:
LOCUS   N58448   599 bp   mRNA   EST   28-JAN-1997
DEFINITION yv71e03.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:248188 3', mRNA sequence.
ACCESSION   N58448
VERSION   N58448.1 GI:1202338
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 599)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE   Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
COMMENT   On Jan 24, 1995 this sequence version replaced gi:634312.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 881   Std Error: 0.00
            Seq primer: m13 -40 forward
            High quality sequence stop: 358.
FEATURES   Location/Qualifiers
     source   1..599
                     /organism="Homo sapiens"
                     /db_xref="GDB:3797434"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:248188"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT   130 a   158 c   152 g   155 t   4 others
ORIGIN

alignment_scores:
    Quality: 81.50   Length: 153
    Ratio: 1.019   Gaps: 4
Percent Similarity: 52.288   Percent Identity: 24.183

alignment_block:
US-09-252-691-7056 x N58448/rev

Align seg 1/1   to reverse of: N58448   from: 1   to: 599
```

```
 27 SerSerArgGlnAlaThrArgArgThrProGlnProThrArgVa 43
    ::::|:::|::||||:||||||| |||   ||  |||
507 ACAGATAAATCAGCCAGGAGAGAGGCCCANTGCCAGTTCTCGCGAAGGCC 458

 43 lIleLeuPheAsnLysProTyrAsp..ValLeuProGlnPheThrAspGl 59
    :::   |||   |  |:: ||  ||
457 AAGGCTAAAGCTGAAGCTATTCGAATCGGTCGCAGCTCTGCACACAACA 408

 59 uAlaGlyArgSerThrLeuLysAspPheIleProValGlnGlyValTyrA 76
    :  |||  ||  ||||::::|  ::|    |||   |||
407 TAATGGAGATGCAGAGCTTCACTGACTGTGGCCAGCAGTATGTCAGCG 358

 76 laAlaGlyArgLeuAspArgAspSerGluGlyLeuLeuValLeuThrAsn 92
    || :|:::|||  ::||||||||:| ::|||||:  ::|||
357 CGTTCTCCAAACTGGCCAAGGACTCCAACACTACTACCGCCCTCCAAC 308

 93 AspGlyValLeuGlnAlaArgLeuThrGlnProGlyLysArgThrGlyLy 109
    |||   :|   ||  || ||  |:|:|||||
307 CCTGGCCATGTCACCAGCATGTGGCCAGGCCATGGGTGTATATGGAGC 258

109 sIleTyrTyrValGlnValGluGlyGluProAspAspAlaSerLeuAlaL 126
    ::|   |||   || |||||||:   |  |:|    ||   |:|
257 CCTCACCAAAGCCCAGTGCCAGGGACTCCAGACTCCACTCTCCAGTGGGA 208

126 ysLeuArgAsnGlyValThrLeuAsnAspGlyProThrLeuProAlaGly 142
    :|:|::||  ::|  :::  ::::|||
207 GCAGCAGAGAT.....GTCCAGGGTACAGATGCAAGTCTTGATGAGGAA 164

143 IleGluArgValAsnGluPro.......GluTrpLeuTrpProArgAsnPr 157
    :|||:||   |||   |||          :||:|||||||
163 CTTGATCGAGTCAAGATGAGTTAGTGAGCTGGAGCTGGCCTTGGCCCAGGGAGTCT 114

157 oProIleArgGluArg...............LysSerIleProT 167
    :::   |||                          ||:|:||
113 GGGGACAAGAAGCAGATTTCCTGATTCTGGCTCAGCTTCCCTGCCAA 64

167 hrSerTrp 169
    |||
 63 GATTTTGG 56
```

OM of: US-09-252-691-7056 to: GenEmbl:*   out_format : pfs

Date: Jun 10, 2000 2:36 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame-p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09252691/runat_05062000_101737_1261/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEX=0.000
-OGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human4o.cd1 -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: GenEmbl:*
Database sequences: 882769
Database length: -48639 5729
Search time (sec): 562.010000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_ba2:AE000213	+	52.00	917.26	4.9e-43	10959	! AE000213 Escherichia coli K-1
gb_bal:D90748	-	52.00	915.15	6.4e-43	15007	! D90748 Escherichia coli genom
gb_bal:D90749	-	52.00	913.13	8.3e-43	20284	! D90749 Escherichia coli genom
gb_htg7:AC022161	+	42.00	714.63	9.5e-32	180993	! AC022161 Homo sapiens chromc
gb_htg6:AC008911	+	28.00	455.78	2.5e-17	226805	! AC008911 Homo sapiens chromc
gb_htg7:AC009127	+	26.00	420.33	2.3e-15	186591	! AC009127 Homo sapiens chromc
gb_ba2:U32752	+	15.00	236.64	4.0e-05	11833	! U32752 Haemophilus influenzae
gb_ba2:SLE250721	+	12.00	186.67	0.0242	5476	! AJ250721 Synechococcus leopoli
gb_bal:D90899	+	12.00	165.22	0.3785	133859	! D90899 Synechocystis sp. PCC
gb_ba2:AE001942	+	10.00	145.87	4.53	9999	! AE001942 Deinococcus radiodura
gb_bal:BACDIA		10.00	138.91	11.06	28206	! L09228 Bacillus subtilis spoV
gb_bal:BSUB0013		10.00	125.17	64.39	218470	! D90726 Bacillus subtilis com
gb_pr3:HSN113A11		9.00	122.01	96.64	22633	! Z83856 Human DNA sequence frc
gb_bal:SHGCPIR		9.00	111.56	369.10	107379	! X86780 S.hygroscopicus gene
gb_htg4:AC011098	+	9.00	108.95	515.58	158329	! AC011098 Homo sapiens clone
gb_pr5:AC004021	+	9.00	108.42	551.92	171370	! AC004021 Human PAC clone RP1
gb_htg12:AC006247		9.00	108.30	560.22	174368	! AC006247 Drosophila melanoga
gb_htg8:AC022255	+	9.00	107.76	600.39	188975	! AC022255 Homo sapiens chromc
gb_htg11:AC021810	+	9.00	107.72	603.79	190218	! AC021810 Homo sapiens chromc
gb_ba2:ECU93212	+	8.00	129.12	38.82	507	! U93212 Escherichia coli scaveng
gb_pl2:AF115188		8.00	127.19	49.73	676	! AF115188 Phaseolus maculatus sm
gb_pl2:AF115195		8.00	127.15	49.98	680	! AF115195 Phaseolus polystachios
gb_ro:AF030343	+	8.00	123.69	77.85	1138	! AF030343 Mus musculus peroxisc
gb_pr4:AF177472	+	8.00	121.81	99.08	1506	! AF177472 Homo sapiens P2X1 pur
gb_ov:GGU22014	+	8.00	121.66	101.01	1540	! U22014 Gallus gallus C-type le
gb_ba2:ECU33213	+	8.00	119.22	138.05	2214	! U33213 Escherichia coli thiol
gb_ro:MMPSYPS	+	8.00	117.79	141.80	2284	! Z83956 M.musculus spermidine s
gb_pr3:HSM801378	+	8.00	117.71	165.96	2742	! AL133100 Homo sapiens mRNA; cl
gb_pl1:SCYNL115C		8.00	117.40	167.62	2774	! Z71391 S.cerevisiae chromosome
gb_bal:RIRSTPSTA		8.00	114.22	174.47	2906	! M31887 R.tsutsugamushi heat sh
gb_in1:CEU66261		8.00	114.22	262.44	4670	! U66261 Caenorhabditis elegans
gb_ba2:AE000230	+	8.00	109.03	510.56	10118	! AE000230 Escherichia coli K-1
gb_ba2:AE002071		8.00	108.86	521.40	10368	! AE002071 Deinococcus radiodur
gb_ba2:AF057718	+	8.00	108.46	548.77	11003	! AF057718 Agrobacterium tumefa
gb_ba2:AE001207		8.00	108.26	563.63	11350	! AE001207 Treponema pallidum s
gb_bal:AB006858	+	8.00	107.82	595.92	12109	! AB006858 Agrobacterium tumefa
gb_bal:D90770		8.00	104.69	890.58	19312	! D90770 E.coli genomic DNA, Kc
gb_bal:D90771		8.00	104.68	891.14	19326	! D90771 E.coli genomic DNA, Kc
gb_pr1:AP000257	-	8.00	101.17	1.4e+03	32609	! AP000257 Homo sapiens genomic

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_in1:CEF14F3	+	8.00	100.71	1.5e+03	34967	! 249937 Caenorhabditis eleg
gb_pl1:SCCXIV38K	+	8.00	100.18	1.6e+03	37800	! 269382 S.cerevisiae 37,800
gb_in2:CELF57C12	+	8.00	99.46	1.7e+03	42080	! U41554 Caenorhabditis eleg
gb_bal:SC51A		8.00	99.39	1.8e+03	42527	! AL121596 Streptomyces coel
gb_in1:CEC10C6	+	8.00	99.20	1.8e+03	43788	! 283217 Caenorhabditis eleg
gb_htg7:AC008879		8.00	96.79	2.5e+03	62697	! AC008879 Homo sapiens chro
gb_htg7:AC018570	+	8.00	96.48	2.6e+03	65588	! AC018570 Homo sapiens chro
gb_pl2:ATAC015985	+	8.00	95.77	2.8e+03	72902	! AC015985 Arabidopsis thali
gb_pt3:HS833B7		8.00	94.62	3.2e+03	86574	! AL008637 Human DNA sequenc
gb_htg7:AC020983	+	8.00	94.53	3.3e+03	87721	! AC020983 Homo sapiens chro
gb_pl2:ATAC010927	+	8.00	93.91	3.5e+03	96232	! AC010927 Arabidopsis thali
gb_pr1:AP000018	+	8.00	93.65	3.7e+03	100000	! AP000018 Homo sapiens gen
gb_pr1:AP000098	+	8.00	93.65	3.7e+03	100000	! AP000098 Homo sapiens gen
gb_pr1:AP000159	+	8.00	93.65	3.7e+03	100000	! AP000159 Homo sapiens gen
gb_pr1:AP000174	+	8.00	93.65	3.7e+03	100000	! AP000174 Homo sapiens gen
gb_in1:CEY52B11A	+	8.00	93.20	3.9e+03	106949	! AL032654 Caenorhabditis e
gb_htg5:AC017146	+	8.00	92.41	4.3e+03	120391	! AC017146 Drosophila melan
gb_pt3:AF042091	+	8.00	92.37	4.3e+03	121019	! AF042091 Homo sapiens chr
gb_htg10:AC012418	+	8.00	92.03	4.5e+03	126010	! AC012418 Homo sapiens clo
gb_pr4:AC004976	+	8.00	91.40	4.9e+03	127425	! AC004976 Homo sapiens PAC
gb_pr4:AC004936	+	8.00	90.74	5.3e+03	139949	! AC004936 Homo sapiens chr
gb_htg3:AC009436	+	8.00	90.72	5.3e+03	154387	! AC009436 Homo sapiens chr
gb_htg7:AC008790	+	8.00	90.59	5.4e+03	154740	! AC008790 Homo sapiens chr
gb_htg5:AC011879	+	8.00	90.31	5.6e+03	157754	! AC011879 Homo sapiens clo
gb_htg3:AC007997	+	8.00	90.29	5.6e+03	164564	! AC007997 Homo sapiens clo
gb_htg9:AC010865	+	8.00	90.23	5.7e+03	165169	! AC010865 Homo sapiens clo
gb_htg2:AC019103	+	8.00	90.18	5.7e+03	166674	! AC019103 Homo sapiens chr
gb_htg2:AL139239	+	8.00	90.10	5.8e+03	167848	! AL139239 Homo sapiens chr
gb_htg7:AC021661	+	8.00	89.81	6.0e+03	169867	! AC021661 Homo sapiens clo
gb_ba2:AE001826		8.00	89.80	6.0e+03	177319	! AE001826 Deinococcus radi
gb_pr4:AC004988	+	8.00	89.48	6.3e+03	177466	! AC004988 Homo sapiens PAC
gb_htg6:AC012621		8.00	89.42	6.3e+03	186141	! AC012621 Homo sapiens chr
gb_htg5:AC018512		8.00	89.34	6.4e+03	187992	! AC018512 Homo sapiens chr
gb_in2:AC008369		8.00	89.26	6.4e+03	190132	! AC008369 Drosophila melan
gb_htg1:AL138964	+	8.00	89.11	6.6e+03	192366	! AL138964 Homo sapiens chr
em_htg1:AC005675		8.00	89.04	6.6e+03	196939	! AC005675 Homo sapiens chr
gb_bal:AB016260		8.00	88.79	6.8e+03	198948	! AB016260 Agrobacterium tu
gb_htg5:AC009109		8.00	88.56	7.0e+03	206479	! AC009109 Homo sapiens chr
gb_htg5:AC009022		8.00	88.35	7.2e+03	213666	! AC009022 Homo sapiens chr
gb_htg4:AC010859		8.00	88.25	7.3e+03	220270	! AC010859 Homo sapiens clo
gb_htg7:AC012325		8.00	88.01	7.6e+03	223594	! AC012325 Homo sapiens chr
gb_htg1:CEY36E3		8.00	87.95	7.6e+03	231928	! 298867 Caenorhabditis ele
gb_ba2:AL022474		8.00	87.49	8.1e+03	233941	! AL022474 Caenorhabditis e
gb_pat:AR032570		7.00	126.84	51.98	250448	! AR032570 Sequence 182 from pa
gb_pat:I29310		7.00	126.84	51.98	46	! I29310 Sequence 182 from pate
gb_pat:I90984		7.00	126.84	51.98	46	! I90984 Sequence 182 from pate
gb_in2:FHU11819		7.00	120.19	122.04	46	! U11819 Fasciola hepatica clo
gb_pr2:S65841S02		7.00	118.91	143.76	124	! S65841 gamma-aminobutyric ac
gb_vi:AF048467		7.00	117.25	177.79	150	! AF048467 HIV-1 patient Q45 1
gb_pl1:KVETS1		7.00	116.49	195.98	192	! X00730 Kluyveromyces lactis
gb_om:AF049887		7.00	116.40	198.33	215	! AF049887 Canis familiaris po
gb_sts:AU028029		7.00	115.81	213.90	218	! AU028029 Rattus norvegicus,
gb_pr1:HS132A7F		7.00	115.45	223.91	238	! 256627 H.sapiens CpG island
gb_pr1:HS15A9F		7.00	114.84	242.22	251	! 257012 H.sapiens CpG island
gb_ro:MMTEG190B		7.00	114.42	255.81	275	! X80422 M.musculus tex190 mRN
gb_pr1:HS155E4F		7.00	114.37	257.31	293	! 256961 H.sapiens CpG island
gb_ov:CHKIGLK		7.00	114.26	261.06	295	! M62984 Chicken Ig light-chai
gb_ov:CHKIGLR		7.00	114.26	261.06	300	! M62967 Chicken Ig light-chai
gb_ov:CHKIGLU		7.00	114.26	261.06	300	! M62968 Chicken Ig light-chai
gb_ov:CHKIGLV		7.00	114.26	261.06	300	! M62969 Chicken Ig light-chai
gb_ov:CHKIGLW		7.00	114.26	261.06	300	! M62970 Chicken Ig light-chai
gb_ov:CHKIGLX		7.00	114.26	261.06	300	! M62971 Chicken Ig light-chai
gb_ov:CHKIGLAG		7.00	114.19	263.31	300	! M21078 Chicken Ig rearranged
gb_sy:AF157755		7.00	114.19	263.31	303	! AF157755 Synthetic construct
gb_sy:AF157763		7.00	114.19	263.31	303	! AF157763 Synthetic construct
gb_sy:AF157767		7.00	114.19	263.31	303	! AF157767 Synthetic construct
gb_ov:CHKIGLAA		7.00	114.06	267.79	303	! M21072 Chicken Ig rearranged
gb_sts:G05174		7.00	113.80	269.28	309	! G05174 human STS WI-9758..10
gb_pr1:HS90A4R		7.00	113.72	276.71	311	! 266349 H.sapiens CpG island
gb_pl2:HVC234520	+	7.00	113.72	279.68	321	! AJ284520 Hordeum vulgare gen
gb_ov:CHKIGLAO	-	7.00	113.64	282.64	325	! M34804 Chicken Ig rearranged

Tue Jun 13 10:29:41 2000   us-09-252-691-7056.oligo.rge

Left block

	gb id	accession / description	rank	score	score	
-	gb_ov:CHKIGLAP	M34805 Chicken Ig rearranged la	332	284.86	113.58	7.00
-	gb_ov:CHKIGLVE	M34803 Chicken Ig germline lamb	335	287.07	113.52	7.00
+	gb_pl1:PS30144E	L38748 Pisolithus tinctorius (F	341	291.49	113.40	7.00
+	gb_sts:G47056	Z15521 1 Zebrafish AB Da	343	292.96	113.36	7.00
+	gb_sts:G08158	G08158 human STS CHLC.GATA4G12.	347	295.90	113.28	7.00
-	gb_v1:AF059712	AF059712 Grapevine yellow speck	366	309.79	112.92	7.00
+	gb_v1:AB028465	AB028465 Grapevine yellow speck	367	310.52	112.78	7.00
+	gb_pr2:HSGLAST9	Z31710 H.sapiens GLAST1 gene fo	374	315.61	112.78	7.00
+	gb_sts:G03711	G03711 human STS WI-282. 3/1995	375	316.34	112.76	7.00
+	gb_bal:RP22465	282465 R.prowazekii genomic DNA	378	318.52	112.71	7.00
+	gb_sts:G50275	G50275 SHGC-79880 Human Homo sa	409	340.87	112.18	7.00
+	gb_bal:MVAC91	X92580 M.vaccae gene for 32 kDa	419	348.04	112.02	7.00
+	gb_ov:CHKIGLA2A	M96968 Gallus Gallus Ig rearran	431	356.60	111.83	7.00
+	gb_ov:CHKIGLA2E	M96973 Gallus gallus Ig rearran	431	356.60	111.83	7.00
+	gb_ro:MUSDHFRA1	M18214 Mouse dihydrofolate redu	434	358.73	111.78	7.00
+	gb_ro:MMCOPS01	G01435 Dm0326 Spradling P-eleme	438	361.58	111.72	7.00
+	gb_pl1:HUMNRAMPO2	AF114236 Mus musculus signalosc	520	419.13	110.57	7.00
-	gb_pl1:GMU15692	L36593 Homo sapiens integral me	525	422.60	110.50	7.00
-	gb_pl1:GMU15692	U15592 Gigaspora margarita puta	533	428.13	110.34	7.00
+	gb_ov:FSALLGF	M57735 M.glutinosa insulin-like	582	461.80	109.81	7.00
-	gb_ov:CHKIGLAVF	M33052 Chicken Ig rearranged li	606	478.15	109.54	7.00
+	gb_pl1:D89939	A46914 Sequence 10 from Patent	613	482.90	109.46	7.00
+	gb_pl2:AF15189	AF15189 Phaseolus maculatus in	613	482.90	109.46	7.00
+	gb_in1:AF001870	AF001870 Rhipicephalus appendic	629	493.72	109.26	7.00
+	gb_pl1:D89939	D89939 Schizosaccharomyces pomb	632	495.75	109.17	7.00
-	gb_pl2:AF15189	X03287 Strongylocentrotus purpu	640	501.15	109.06	7.00
+	gb_in1:SPSPEC11	294004 Unidentified bacterium D	651	508.55	109.03	7.00
-	gb_bal:UB294004	AR035064 Sequence 73 from paten	654	510.57	109.03	7.00
-	gb_pat:AR035064	AR035065 Sequence 75 from paten	654	510.57	109.03	7.00
-	gb_pat:AR035065	AL16915 Botrytis cinerea stral	660	514.60	108.97	7.00
-	gb_pat:CNS01DJV	AF15247 Phaseolus xolocotzli i	663	516.61	108.94	7.00
+	gb_pl2:AF15247	A43751 Sequence 7 from Patent W	667	519.29	108.90	7.00
+	gb_pat:A43751	AF15187 Phaseolus maculatus sm	675	524.65	108.82	7.00
+	gb_pl2:AF15187	AF15191 Phaseolus juquilensis	676	525.32	108.81	7.00
+	gb_pl2:AF15191	AF15192 Phaseolus juquilensis	676	525.32	108.81	7.00
+	gb_pl2:AF15192	AF15192 Phaseolus juquilensis	677	525.99	108.80	7.00
+	gb_pl2:AF15182	AF15182 Phaseolus salicifolius	677	525.99	108.80	7.00
-	gb_pl2:AF15193	AF15193 Phaseolus smilacifoliu	677	525.99	108.80	7.00
-	gb_pl2:AF15194	AF15194 Phaseolus sinuatus sma	678	526.65	108.79	7.00
-	gb_pl2:AF15196	AF15196 Phaseolus polystachios	678	526.65	108.79	7.00
-	gb_pl2:AF15190	AF15190 Phaseolus jaliscanus s	678	526.65	108.70	7.00
-	gb_pl2:AF15199	AF15199 Phaseolus jaliscanus s	697	539.33	108.60	7.00
+	gb_pat:E17140	E17140 Human mRNA for type II m	697	539.33	108.60	7.00
-	gb_pri:AB015628	AB015628 Homo sapiens mRNA for	709	547.32	108.49	7.00
+	gb_pat:HSPR264	X62447 H.sapiens PR264 mRNA. 4/	719	553.95	108.38	7.00
-	gb_pl1:YSCRPR1C	L12746 Saccharomyces carlsberge	720	554.62	108.30	7.00
-	gb_pl2:CNS01ENU	AL14209 Botrytis cinerea seri	720	554.62	108.30	7.00
-	gb_pl1:HSHSP70P	X13229 Human hsp70B gene 5' reg	729	560.58	108.20	7.00
+	gb_pr3:HSGLRA2G5	AF053491 Homo sapiens glycine I	729	560.58	108.11	7.00
+	gb_sy:AF157780	AF157780 Synthetic construct cl	740	567.85	108.03	7.00
-	gb_pl1:HSAICL	X96719 H.sapiens mRNA for AICL	750	574.45	107.79	7.00
+	gb_in1:AAU52972	U52972 Aurelia aurita 18S ribos	759	580.38	107.77	7.00
+	gb_pat:A82998	A82998 Sequence 5 from Patent W	759	580.38	107.48	7.00
+	gb_pl2:MAZ99983	Z99983 Musa acuminata mRNA for	789	600.07	107.48	7.00
+	gb_v1:EVI7935	AL14079 Anopheles gambiae STS	823	622.26	107.31	7.00
-	gb_v1:RVU07763	AJ007935 Ectromelia virus for	823	622.26	107.06	7.00
+	gb_sts:CNS01KTU	U07763 Rabbitpox virus Utrecht	845	636.55	106.67	7.00
-	gb_pat:A79811	AL144431 Anopheles gambiae STS	877	657.24	106.38	7.00
-	gb_pat:A79812	AL149075 Anopheles gambiae STS	929	690.65	106.21	7.00
-	gb_pat:S83210	AL148549 Anopheles gambiae STS	949	703.43	106.04	7.00
+	gb_in2:AF145277	X87402 K.lactis DNA for externa	971	717.44	105.97	7.00
-	gb_ba2:AF074437	X16162 Human DNA homologous to	973	718.71	105.93	7.00
-	gb_sma:SMA243121	U51965 Dactylella huisuniana in	995	732.68	105.82	7.00
+	gb_v1:EVI244012	AL140779 Anopheles gambiae STS	1020	748.50	105.79	7.00
-	gb_v1:D86537	AJ007935 Ectromelia virus for	1032	756.07	105.78	7.00
		L36536 Human (clone cHBC2) E-ca	1035	757.96	105.40	7.00
		AF219903 Botrytis cinerea seri	1038	759.85	105.39	7.00
		U07763 Rabbitpox virus Utrecht	1055	770.55	105.24	7.00
		AL144431 Anopheles gambiae STS	1059	773.06	105.24	7.00
		A79811 Sequence 1 from Patent	1059	773.06		7.00
		A79812 Sequence 2 from Patent	1061	774.32		7.00
		S83210 heat shock protein 70-a	1070	779.97		7.00
		AF145277 Trichomonas vaginalis	1123	813.11		7.00
		AF074437 Serratia marcescens A	1124	813.73		7.00
		AJ243121 Serratia marcescens I	1137	821.83		7.00
		AJ244012 Ectromelia virus prov	1150	829.91		7.00
		D86537 Hepatitis C virus gene				

Right block

rank	score	score			gb id	accession / description
1167	840.46	105.14	7.00	-	gb_ov:U71229	U71229 Desmognathus quadram
1170	842.32	105.12	7.00	-	gb_ov:AF077819	AF077819 Cyprinus carpio hy
1226	876.90	104.81	7.00	+	gb_pr2:HSPFKR089	X16917 Human PFKL gene for
1269	903.31	104.55	7.00	+	gb_v1:SINE2GP	M13818 Sindbis virus E2 gly
1275	906.99	104.52	7.00	+	gb_v1:VACB13R	D00581 Vaccinia virus gene
1280	910.05	104.52	7.00	+	gb_bal:RHMNIFHW	K01620 Rhizobium japonicum
1281	910.66	104.30	7.00	+	gb_pat:E00713	E00713 Genomic DNA encoding
1322	935.69	104.30	7.00	+	gb_pat:AR035496	AR035496 Sequence 5 from pa
1322	935.69	104.30	7.00	+	gb_pat:I59447	I59447 Sequence 5 from pate
1322	935.69	104.19	7.00	+	gb_prl:HUMCRABPO1	M97814 Human retinoic acid
1345	949.69	104.10	7.00	+	gb_bal:RLROPB	X80767 R.leguminosarum ropB
1363	960.62	104.04	7.00	+	gb_ro:MMSHP1S05	U65995 Mus musculus SH2 pho
1376	968.50	104.03	7.00	+	gb_ro:MMMDR2X	Z35287 M.musculus (129/OLA)
1376	969.10	103.94	7.00	+	gb_pat:A96836	A96836 Sequence 869 from Pa
1377	980.60	103.82	7.00	+	gb_ov:CHKELA	M18633 Chicken elastin mRNA
1396	993.89	103.82	7.00	+	gb_bal:BLA243948	AJ243948 Bifidobacterium la
1418	995.70	103.80	7.00	+	gb_pat:E01495	E01495 cDNA encoding tropoe
1421	998.11	103.78	7.00	+	gb_pat:E01496	E01496 DNA encoding tropoel
1421	1.0e+03	103.75	7.00	+	gb_pl1:ASNOMT	L22091 Aspergillus parastie
1425	1.0e+03	103.60	7.00	+	gb_ov:CHKXPACA	D31896 Gallus gallus mRNA f
1429	1.0e+03	103.57	7.00	+	gb_v1:CPVWPV	M74803 Trypanosoma congolen
1436	1.0e+03	103.54	7.00	+	gb_v1:VACSP12A	M14217 Cowpox virus slike-p
1468	1.0e+03	103.51	7.00	+	gb_bal:ULO240981	M24218 Vaccinia virus serin
1475	1.1e+03	103.29	7.00	+	gb_prl:HUMKCHA	J04204 Bos taurus 32 kd ace
1482	1.1e+03	103.25	7.00	+	gb_om:BOV39KD	D43763 Chicken mRNA for 3be
1488	1.1e+03	103.23	7.00	+	gb_ov:CHKD3BHD5B	U38240 Oryctolagus cuniculu
1538	1.1e+03	103.23	7.00	+	gb_om:OCU38240	L02750 Human potassium chan
1547	1.1e+03	103.07	7.00	+	gb_in2:AF150875	AF150875 Giardia intestinal
1551	1.1e+03	103.05	7.00	+	gb_om:AF064455	AF064455 Canis familiaris m
1551	1.1e+03	103.04	7.00	+	gb_ro:MUSHSD3	L41519 Mus musculus 3-ketos
1589	1.1e+03	103.02	7.00	+	gb_v1:CCPPOL	M21938 Cricket paralysis v1
1592	1.1e+03	103.01	7.00	+	gb_ro:RNMKP3	X94185 R.norvegicus mRNA fo
1594	1.1e+03	103.00	7.00	+	gb_ba2:PLMPTAR	M21299 Agrobacterium tumefa
1596	1.2e+03	102.89	7.00	+	gb_ro:HAMORF	L38709 Mesocricetus auratus
1600	1.2e+03	102.83	7.00	+	gb_ro:MUSKEFOSTE	L16919 Mus musculus (pK4l1A
1603	1.2e+03	102.83	7.00	+	gb_pat:I77057	I77057 Sequence 4 from pate
1604	1.2e+03	102.80	7.00	+	gb_pat:A59209	A59209 Sequence 3 from Pate
1631	1.2e+03	102.70	7.00	+	gb_pll:RNAMELIN2	Z50084 R.norvegicus mRNA fo
1648	1.2e+03	102.61	7.00	+	gb_pll:ANDPIGEN	X89449 A.niger mRNA for dis
1648	1.2e+03	102.57	7.00	+	gb_bal:SSAEF1A	X70701 S.solfaricus gene
1653	1.2e+03	102.46	7.00	+	gb_ba2:SMU81967	U81967 Serratia marcescens
1680	1.2e+03	102.46	7.00	+	gb_in1:AB024086	AB024086 Caenorhabditis ele
1701	1.2e+03	102.46	7.00	+	gb_pat:A70007	A70007 Sequence 2 from Pate
1711	1.3e+03	102.31	7.00	+	gb_ro:AF017636	AF017636 Mesocricetus aurat
1711	1.3e+03	102.30	7.00	+	gb_ph:BPHINTXIS	M34832 Bacteriophage phi-1
1740	1.3e+03	102.16	7.00	+	gb_om:CFU08596	U08596 Canis familiaris del
1740	1.3e+03	102.10	7.00	+	gb_prl:HUMPCD	A19151 ovine IL-1 alpha. 5/
1741	1.3e+03	101.94	7.00	+	gb_prl:HUMPCH	M55515 Human potassium chan
1780	1.3e+03	101.91	7.00	+	gb_prl:HUMSC35A	L23499 Homo sapiens voltage
1781	1.3e+03	101.83	7.00	+	gb_prl:PAPOBR	M90104 Human splicing facto
1820	1.3e+03	101.83	7.00	+	gb_bal:PAPOBR	X76994 P.aeruginosa pobR an
1836	1.3e+03	101.82	7.00	+	gb_pr3:AF048712	AF048712 Homo sapiens Kv4.3
1879	1.3e+03	101.73	7.00	+	gb_pr3:AF205856	AF205856 Homo sapiens potas
1890	1.3e+03	101.71	7.00	+	gb_pr3:HSM800564	Z49289 S.cerevisiae chromos
1911	1.3e+03	101.69	7.00	+	gb_pr5:AF161528	U22063 Saccharomyces cerevi
1911	1.3e+03	101.63	7.00	+	gb_pat:A59207	X73570 S.cerevisiae YTA2 ge
1913	1.3e+03	101.63	7.00	+	gb_ro:RATHSDI	A79820 Sequence 10 from Pat
1914	1.3e+03	101.63	7.00	+	gb_v1:AB000139	U94352 Human manic fringe p
1939	1.4e+03	101.61	7.00	+	gb_om:AF198445	U43153 Saccharomyces cerevi
1939	1.4e+03	101.52	7.00	+	gb_pr3:AF048713	A85166 Saccharomyces cerevi
1939	1.4e+03	101.47	7.00	+	gb_pr4:AF205857	A85164 Sequence 3 from Pate
1947	1.4e+03	101.43	7.00	+	gb_pr3:SCYJL014W	A85168 Sequence 1 from Pate
1952	1.4e+03	101.39	7.00	+	gb_pll:SCU22063	U42627 Rattus norvegicus du
1968	1.4e+03	101.36	7.00	+	gb_pll:SCYTA2A	
1968	1.4e+03	101.29	7.00	+	gb_pat:A79820	
1968	1.4e+03	101.19	7.00	+	gb_pr3:HSU94352	
1975	1.4e+03	101.19	7.00	+	gb_pr3:SCU43153	
2002		101.19	7.00	+	gb_pat:A85166	
2016			7.00	+	gb_pat:A85164	
2028			7.00	+	gb_pat:A85168	
2042			7.00	+	gb_ro:RNU42627	
2049						
2072						
2104						
2104						
2104						

±	Name (L)	#	Accession / Description (L)	E	Score	cut	#	E	%	cut	±	Name (R)	Accession / Description (R)
-	gb_om:BOVRP3A	2135	D13613 Bovine mRNA for rabphil	1.4e+03	101.09	7.00	3139	2.0e+03	98.50	7.00	+	gb_prl:HSA005898	AJ005898 Homo sapiens mRNA
-	gb_pat:E07809	2135	E07809 cDNA encoding rabphilin	1.4e+03	101.09	7.00	3140	2.0e+03	98.50	7.00	-	em_in:DM14DC10Z	L32644 Drosophila melanogas
+	gb_pll:YSCTCP20	2191	L27698 Yeast chaperonin protei	1.4e+03	100.91	7.00	3145	2.0e+03	98.49	7.00	-	gb_pll:CUCACO	D29629 Pumpkin mRNA for aco
-	gb_bal:SSU33211	2196	U33211 Synechococcus sp. FGAM	1.4e+03	100.90	7.00	3212	2.0e+03	98.35	7.00	-	gb_inl:HSDGII	X56807 Human DSC2 mRNA for
.	gb_ov:S76880	2202	S76880 XIdx-bHLH function domi	1.5e+03	100.88	7.00	3266	2.0e+03	98.23	7.00	+	gb_inl:DMRTGENE	X95956 D.melanogaster mRNA
+	gb_pll:CHHSP70	2203	X81860 C.herbarum mRNA for hea	1.5e+03	100.88	7.00	3279	2.1e+03	98.21	7.00	+	gb_ro:RNTRHRMR	X64630 R.norvegicus mRNA fo
+	gb_pat:A76755	2206	A76755 Sequence 1 from Patent	1.5e+03	100.87	7.00	3320	2.1e+03	98.12	7.00	+	gb_pat:A85269	A85269 Sequence 1 from Pate
+	gb_pat:A76760	2206	A76760 Sequence 1 from Patent	1.5e+03	100.87	7.00	3336	2.1e+03	98.05	7.00	+	gb_ro:NCR9759	AJ009759 Neurospora crassa
+	gb_prl:HS90KTAA	2215	X79089 H.sapiens mRNA for 90K	1.5e+03	100.84	7.00	3359	2.1e+03	98.04	7.00	-	gb_bal:MTXACS	M63968 M.soehngenii acetyl-
+	gb_bal:CYAPETC	2221	J03855 C.nostoc Rieske iron-su	1.5e+03	100.82	7.00	3363	2.1e+03	98.04	7.00	+	gb_ro:RNPTHYRHR	X66726 R.norvegicus mRNA fo
-	gb_pr3:HSM800400	2222	AL050074 Homo sapiens mRNA; cD	1.5e+03	100.82	7.00	3364	2.1e+03	97.91	7.00	-	gb_ro:S60053	S60053 thyrotropin-releasi
+	gb_ov:AF014368	2225	AF014368 Danio rerio transcrip	1.5e+03	100.81	7.00	3428	2.1e+03	97.87	7.00	-	gb_bal:SRPSAEF	X76767 S.solfataricus rps12
-	gb_ba2:RHMACTS	2267	L39938 Sinorhizobium meliloti	1.5e+03	100.69	7.00	3447	2.1e+03	97.85	7.00	-	gb_prl:HUM8DC21z	L35676 Homo sapiens (subclo
+	gb_prl:HSU18247	2275	U18247 Human p67 mRNA, complet	1.5e+03	100.66	7.00	3460	2.1e+03	97.80	7.00	-	gb_pr3:HSU94333	U94333 Human C1q/MBL/SPA re
+	gb_pr3:HSM802005	2285	AL133326 Homo sapiens mRNA; cD	1.5e+03	100.63	7.00	3485	2.2e+03	97.68	7.00	-	gb_pr5:AF184971	AF184971 Homo sapiens class
+	gb_prl:HSGAT1MR	2298	X54673 H.sapiens GAT1 mRNA for	1.5e+03	100.59	7.00	3546	2.2e+03	97.60	7.00	+	gb_ro:RNU58279	U58279 Rattus norvegicus M1
+	gb_pll:SCYOL148C	2299	Z74490 S.cerevisiae chromosome	1.5e+03	100.54	7.00	3591	2.2e+03	97.56	7.00	+	gb_v1:PS2BETACD	M81486 Poa semilatent virus
+	gb_bal:TFERBCO	2317	D90113 Thiobacillus ferrooxida	1.5e+03	100.52	7.00	3610	2.2e+03	97.54	7.00	+	gb_inl:BPAMPHIHH	Y13858 Branchiostoma florid
+	gb_om:RABMM9	2323	D26514 Rabbit mRNA for matrix	1.5e+03	100.50	7.00	3624	2.2e+03	97.49	7.00	-	gb_bal:ECTREMER2	Y10102 Escherichia coli par
+	gb_bal:ECPAEY	2329	Y09828 E.chrysanthemi paey gen	1.5e+03	100.43	7.00	3648	2.3e+03	97.42	7.00	-	gb_pr4:AF132609	AF132609 Homo sapiens histo
+	gb_v1:MEAFUS	2355	AF014915 Measles virus, membrane	1.5e+03	100.38	7.00	3687	2.3e+03	97.39	7.00	-	gb_ro:RATTRHRA	D17469 Rat mRNA for thyrotr
+	gb_pl2:AF118038	2372	AF118038 Juglans regia maturas	1.5e+03	100.35	7.00	3702	2.3e+03	97.32	7.00	-	gb_ro:MDSIRS1X	L24563 Mouse insulin recept
+	gb_v1:AF179432	2373	AF179432 Measles virus strain	1.6e+03	100.35	7.00	3741	2.3e+03	97.30	7.00	+	gb_pl2:AF145300	AF145300 Arabidopsis thalia
+	gb_v1:MEV133108	2383	AJ133108 Measles virus mRNA fc	1.6e+03	100.31	7.00	3756	2.3e+03	97.11	7.00	+	gb_prl:AGMPRIRRS1	D64157 Cercopithecus aethio
-	gb_v1:MEAFP	2384	D00090 Measles virus mRNA for	1.6e+03	100.30	7.00	3862	2.4e+03	97.09	7.00	-	gb_pr4:AF076952	AF076952 Homo sapiens xerod
+	gb_v1:PAMVFUS	2384	X05597 Measles virus (Halle) m	1.6e+03	100.24	7.00	3871	2.4e+03	97.09	7.00	-	gb_prl:HSBMPRII	Z48923 H.sapiens mRNA for B
+	gb_pr3:HUMKCHN	2397	X98217 Homo sapiens voltage-ga	1.6e+03	100.11	7.00	3873	2.4e+03	96.98	7.00	-	gb_ov:PWFGFR	X59380 P.waltlii mRNA for f
+	gb_ro:MFXR1H2	2400	AF124387 Homo sapiens fragile-	1.6e+03	100.04	7.00	3936	2.4e+03	96.97	7.00	-	gb_ov:AF156539	AF156539 Homo sapiens xerod
+	gb_ro:MUSNFKB	2424	M61909 Mouse transcription fac	1.6e+03	100.03	7.00	3943	2.4e+03	96.87	7.00	-	gb_ov:GGY14271	Y14271 Gallus gallus mRNA f
+	gb_prl:ATHAMYB	2469	X98476 Arabidopsis thaliana mR	1.6e+03	99.98	7.00	4000	2.4e+03	96.83	7.00	-	gb_om:D85682	D85682 Bos taurus mRNA for
+	gb_pr2:HSVASPEX1	2494	X98797 H.sapiens VASP gene. 4/19	1.6e+03	99.95	7.00	4027	2.5e+03	96.76	7.00	-	gb_pl2:XCU94336	U94336 Xanthomonas campestr
+	gb_pll:ANPDIAGEN	2500	X97019 A.niger pdiA gene. 4/19	1.6e+03	99.82	7.00	4070	2.5e+03	96.76	7.00	-	gb_v1:ALSIND	V00073 Sindbis virus sequen
+	gb_pl2:ATHPARA	2500	L43081 Arabidopsis thaliana pE	1.6e+03	99.78	7.00	4071	2.5e+03	96.74	7.00	-	gb_v1:TOSINR	V01403 Sindbis virus intrac
+	gb_bal:ECU45854	2529	U45854 Erwinia chrysanthemi ph	1.6e+03	99.78	7.00	4078	2.5e+03	96.71	7.00	+	gb_prl:HSA011972	AB020708 Homo sapiens mRNA
+	gb_ro:MUSMDR2A	2529	M74151 Mouse P-glycoprotein (r	1.6e+03	99.71	7.00	4099	2.5e+03	96.67	7.00	+	gb_om:OATCRCDEL	AJ011972 Homo sapiens mRNA
+	gb_pat:A45787	2579	A45787 Sequence 25 from Patent	1.7e+03	99.64	7.00	4126	2.5e+03	96.64	7.00	+	gb_pr3:HSITGAD06	U40279 Human beta-2 integri
+	gb_pll:SCBIN2	2594	X82840 Saccharomyces cerevisia	1.7e+03	99.60	7.00	4143	2.5e+03	96.60	7.00	+	gb_ba2:AF010149	AF010149 Pseudomonas aerugi
-	gb_ro:RATNFM	2594	M18628 Rat NF-M middle molecul	1.7e+03	99.54	7.00	4168	2.5e+03	96.59	7.00	-	gb_pl3:AF120494	AF120494 Arabidopsis thalia
-	gb_ov:CHK3BHD54A	2621	D43762 Chicken mRNA for 3beta	1.7e+03	99.51	7.00	4176	2.5e+03	96.41	7.00	-	gb_ba1:AF144078	AF144078 Arabidopsis thalia
-	gb_pl2:AF120491	2650	AF120491 Chicken mRNA for 3beta	1.7e+03	99.49	7.00	4288	2.6e+03	96.31	7.00	+	gb_ba2:AF131202	AF131202 Pantoea citrea 2-k
-	gb_pat:E12795	2665	M83254 Human cardiac potassium	1.7e+03	99.47	7.00	4349	2.6e+03	96.25	7.00	+	gb_pl2:AF119670	AF119670 Magnaporthe grisea
-	gb_prl:HUMPOTCH	2688	X70119 P.putida exbD an	1.7e+03	99.46	7.00	4388	2.6e+03	96.16	7.00	+	gb_pll:AB004547	AB004547 Rhizopus oligospor
-	gb_prl:HUM26SPSP	2701	U67930 Escherichia coli ChuA g	1.7e+03	99.38	7.00	4449	2.7e+03	96.14	7.00	+	gb_pll:TORCPS	L08965 Trichosporon cutaneu
+	gb_bal:AF027979	2709	AF035922 Pseudomonas aeruginos	1.8e+03	99.26	7.00	4463	2.7e+03	96.02	7.00	+	gb_ba2:AC014570	AC014570 Pseudomonas melano
+	gb_pll:MTCYW318	2717	Z97015 Mycobacterium tubercul	1.8e+03	99.25	7.00	4541	2.7e+03	95.85	7.00	-	gb_htg4:AC012884	AC012884 Drosophila melanog
+	gb_pll:ASNOMT1A	2717	L25836 Aspergillus Flavus O-me	1.8e+03	99.20	7.00	4661	2.8e+03	95.83	7.00	-	gb_bal:SCMETALLO	Z11929 S.coelicolor genes f
+	gb_pll:ASNOMT1C	2722	L25834 Aspergillus Flavus O-me	1.8e+03	99.14	7.00	4673	2.8e+03	95.80	7.00	+	gb_om:SSP101	Y10742 S.scrofa mRNA for pl
+	gb_pll:HSU12596	2755	U12596 Human tumor necrosis fa	1.8e+03	99.12	7.00	4692	2.8e+03	95.80	7.00	+	gb_pat:AR026672	AR026672 Sequence 4 from pate
+	gb_pl2:AF087483	2803	AF087483 Arabidopsis thaliana	1.8e+03	99.11	7.00	4692	2.8e+03	95.80	7.00	+	gb_pat:AR026678	AR026678 Sequence 14 from Pat
+	gb_pr2:HSRNA511	2809	X86446 H.sapiens mRNA for 55.1	1.8e+03	98.99	7.00	4692	2.8e+03	95.80	7.00	+	gb_pat:AR029061	AR029061 Sequence 1 from pa
-	gb_prl:HUMPCA	2828	M55513 Human potassium channel	1.8e+03	98.98	7.00	4692	2.8e+03	95.80	7.00	+	gb_pat:AR033965	AR033965 Sequence 1 from pa
+	gb_pl2:NCU23425	2852	U23425 Neurospora crassa ropy	1.8e+03	98.98	7.00	4716	2.8e+03	95.77	7.00	-	gb_pll:ATHBA	D43783 Mouse-ear-cress DNA
-	gb_pat:E12795	2863	E12795 cDNA encoding human 26S	1.9e+03	98.88	7.00	4716	2.8e+03	95.77	7.00	+	gb_pl2:S77076	S77076 beta-amylase [Arabid
-	gb_prl:HUM26SPSP	2867	D78151 Human mRNA for 26S prot	1.9e+03	98.87	7.00	4854	2.9e+03	95.58	7.00	-	gb_pl3:HSM02148	AL137696 Homo sapiens mRNA;
+	gb_bal:AF027979	2918	AF027979 Magnaporthe grisea ca	1.9e+03	98.86	7.00	5062	3.0e+03	95.29	7.00	-	gb_ro:AF132986	AF132986 Mus musculus domes
+	gb_pll:MTCYW318	2922	Z97015 Mycobacterium tubercul	1.9e+03	98.84	7.00	5125	3.0e+03	95.21	7.00	+	gb_pat:I40454	I40454 Sequence 4 from pate
+	gb_pll:ASNOMT1A	2922	L25836 Aspergillus niger subtl	1.9e+03	98.82	7.00	5150	3.0e+03	95.18	7.00	+	gb_pat:A62729	A62729 Sequence 14 from Pat
+	gb_pll:ASNOMT1C	2966	L25834 Aspergillus Flavus O-me	1.9e+03	98.82	7.00	5246	3.1e+03	95.05	7.00	+	gb_ro:MMIRS1	X69722 M.musculus mRNA for
+	gb_bal:AF226283	2969	AF226283 Xanthomonas oryzae pv	1.9e+03	98.82	7.00	5365	3.1e+03	94.90	7.00	+	gb_ro:RNIRS1IRM	X58375 R.Norvegicus IRS-1 m
+	gb_bal:APY15785	2976	Y15785 Aquifex Pyrophilus 23S	1.9e+03	98.80	7.00	5405	3.1e+03	94.85	7.00	-	gb_ro:MMU73199	U73199 Mus musculus Rho-gua
+	gb_pat:A78620	2986	A78620 Sequence 6 from Patent	1.9e+03	98.79	7.00	5413	3.2e+03	94.84	7.00	-	gb_pll:SCYKR031C	Z28256 S.cerevisiae chromos
+	gb_pat:AR064177	2993	AR064177 Sequence 6 from Patent	1.9e+03	98.65	7.00	5455	3.2e+03	94.79	7.00	-	gb_pll:ATPRLDNA	X82824 A.thaliana PRL1 gene
+	gb_ro:RNNFMMR	2993	L19059 R.norvegicus mRNA for n	1.9e+03	98.60	7.00	5471	3.2e+03	94.77	7.00	+	gb_bal:PFL250853	AJ250853 Pseudomonas putida
+	gb_prl:HUMKCHAN	3004	M85217 H.sapiens K+ channel pr	1.9e+03	98.58	7.00	5480	3.2e+03	94.76	7.00	+	gb_ov:CHKIGLVJC	M24403 Chicken germline Ig
+	gb_bal:TN3926PA	3070	X14236 Escherichia coli Plasmi	1.9e+03	98.53	7.00	5509	3.2e+03	94.73	7.00	+	gb_prl:AB011138	AB011138 Homo sapiens mRNA
+	gb_pll:SCSLY1	3091	Z12152 Yeast SLY1 gene. 3/1991	2.0e+03	98.52	7.00	5526	3.2e+03	94.71	7.00	+	gb_om:CFU50709	U50709 Canis familiaris BRC
+	gb_pl3:AF171697	3104	X54323 Yeast SLY1 gene. 3/1991	2.0e+03	—	7.00	5562	3.2e+03	94.71	7.00	+	gb_ro:MMD39818	V39818 Mus musculus tuberin
+	gb_in2:AF177540	3125	AF171697 Sequence SLY1 gene 3/1991	2.0e+03	—	7.00	5637	3.3e+03	94.57	7.00	-	gb_ro:MMU73199	U37775 Mus musculus tuberin
+	gb_ba2:AF139182	3127	AF139182 Bartonella henselae f	2.0e+03	—	7.00	5637	3.3e+03	94.57	7.00	-	gb_ro:RNEPIG	X59831 Rat gene for epididy
+	gb_ba2:PWU88907	3130	U88907 Pseudomonas wisconsinen	2.0e+03	—	7.00	5661	3.3e+03	94.54	7.00	-	gb_ba2:AF106002	AF106002 Pseudomonas putida
							5678	3.3e+03	94.52	7.00	-	gb_ba2:AF007813	AF007813 Leptospira interro

Left column

strand	gb code	pos	match	7.00	%	score
-	gb_pll:YSCSPO14G	5778	! L46807 Saccharomyces cerevisia	7.00	94.41	3.3e+03
-	gb_pr2:S62539	5828	! S62539 insulin receptor substr	7.00	94.35	3.4e+03
+	gb_pll:EMENIRA	5835	! M68900 A.nidulans NIRA protein	7.00	94.34	3.4e+03
+	gb_htg6:AC020186	5891	! AC020186 Drosophila melanogast	7.00	94.28	3.4e+03
+	gb_pat:AR064200	5924	! AR064200 Sequence 6 from paten	7.00	94.24	3.4e+03
+	gb_pat:I89172	5924	! I89172 Sequence 6 from patent	7.00	94.24	3.4e+03
-	gb_htg4:AB033082	5938	! AB033082 Homo sapiens mRNA for	7.00	94.22	3.4e+03
+	gb_pat:AR050631	6152	! AR050631 Sequence 1 from paten	7.00	93.99	3.5e+03
+	gb_pr2:S85963	6152	! S85963 hIRS-1=rat insulin rece	7.00	93.99	3.5e+03
+	gb_pr2:HSNFM	6236	! Y00067 Human gene for neurofil	7.00	93.89	3.6e+03
+	gb_prl:D86985	6239	! D86985 Homo sapiens mRNA for K	7.00	93.89	3.6e+03
+	gb_sy:AF174131	6333	! AF174131 Cloning vector pRS411	7.00	93.79	3.6e+03
+	gb_ro:RATQLIKEA	6389	! L16012 Rattus norvegicus Q-lik	7.00	93.73	3.6e+03
+	gb_ro:RNIGMPOR	6942	! Y00826 Rat mRNA for integral m	7.00	93.17	3.9e+03
+	gb_ro:AF113751	6998	! AF113751 Mus musculus nuclear	7.00	93.12	3.9e+03
+	gb_ov:AB007931	7150	! AB007931 Homo sapiens mRNA for	7.00	92.97	4.0e+03
+	gb_ov:AF130854	7157	! AF130854 Xiphophorus maculatus	7.00	92.97	4.0e+03
+	gb_sy:AF174132	7168	! AF174132 Cloning vector pRS421	7.00	92.96	4.0e+03
+	gb_vi:HS1ATIF1	7241	! M15621 HSV1 (strain F) alpha-t	7.00	92.89	4.0e+03
+	gb_bal:ECMERTREL	8012	! Y09025 Enterobacter cloacae DN	7.00	92.21	4.4e+03
+	gb_pl2:ACTPOLYGAL	8062	! L12019 Actinidia deliciosa var	7.00	92.17	4.4e+03
+	gb_sy:AF174133	8182	! AF174133 Cloning vector pRS421	7.00	92.07	4.5e+03
+	gb_vi:AB003289	8953	! AB003289 Hepatitis GB virus C	7.00	91.47	4.9e+03
+	gb_ba2:U39721	9060	! U39721 Mycoplasma genitalium s	7.00	91.39	4.9e+03
+	gb_ro:MMU271753	9099	! AJ271753 Mus musculus Cx32 gen	7.00	91.36	4.9e+03
+	gb_vi:X16568	9207	! X16568 Measles virus (patient	7.00	91.28	5.0e+03
+	gb_vi:MEAIP3CA	9208	! X16566 Measles virus (strain I	7.00	91.28	5.0e+03
+	gb_sy:AF155773	9470	! AF155773 Gibberella fujikuroi	7.00	91.05	5.1e+03
+	gb_vi:HIV2MDS	9525	! Z48731 Human immunodeficiency	7.00	91.05	5.1e+03
+	gb_bal:PDE010260	9631	! AJ010260 Paracoccus denitrific	7.00	90.98	5.2e+03
+	gb_htg4:AC014284	9692	! AC014284 Drosophila melanogas	7.00	90.94	5.2e+03
+	gb_htg8:AC016612	10063	! AC016612 Homo sapiens chromos	7.00	90.68	5.4e+03
+	gb_ba2:AE001198	10074	! AE001198 Treponema pallidum s	7.00	90.68	5.4e+03
+	gb_bal:SDG64516	10090	! U64516 Shigella dysenteriae s	7.00	90.67	5.4e+03
+	gb_ba1:AE000998	10279	! AE000998 Archaeoglobus fulgid	7.00	90.54	5.5e+03
+	gb_pl2:AF035536	10328	! AF035536 Spinacia oleracea cc	7.00	90.51	5.5e+03
+	gb_ba2:AE000416	10348	! AE000416 Escherichia coli K-1	7.00	90.50	5.5e+03
+	gb_ba1:AE000996	10391	! AE000996 Archaeoglobus fulgid	7.00	90.47	5.5e+03
+	gb_bal:PDENQOURF	10645	! L02354 Paracoccus denitrifica	7.00	90.28	5.7e+03
+	gb_bal:AE000601	10663	! AE000601 Helicobacter pylori	7.00	90.28	5.7e+03
+	gb_ba1:AE000889	10688	! AE000889 Methanobacterium the	7.00	90.22	5.7e+03
+	gb_ba2:AE001290	10784	! AE001290 Chlamydia trachomati	7.00	90.19	5.7e+03
+	gb_in2:LSU54556	10792	! U54556 Litomosoides sigmodont	7.00	90.04	5.8e+03
+	gb_pr4:AF186772	10828	! AF186772 Homo sapiens surfeit	7.00	89.67	5.9e+03
+	gb_pr2:HSPTEL	11080	! Z84812 Human DNA sequence fro	7.00	89.67	5.9e+03
+	gb_ba2:AE000429	11156	! AE000429 Escherichia coli K-1	7.00	89.63	6.1e+03
+	gb_ba2:AE001455	11191	! AE001455 Helicobacter pylori,	7.00	89.59	6.1e+03
+	gb_pl1:SCXYLORFS	11375	! U82209 Homo sapiens chromosom	7.00	89.58	6.2e+03
+	gb_htg4:AC014620	11456	! AC014620 Drosophila melanogas	7.00	89.44	6.3e+03
+	gb_ba2:AE001982	11591	! AE001982 Deinococcus radiodur	7.00	89.36	6.4e+03
+	gb_ba2:AE002001	11703	! AE002001 Deinococcus radiodur	7.00	89.30	6.4e+03
+	gb_ba2:AE002039	11703	! AE002039 Deinococcus radiodur	7.00	89.26	6.5e+03
+	gb_ba1:AE000808	11708	! AE000808 Methanobacterium the	7.00	89.05	6.6e+03
+	gb_ba2:AE000211	11767	! AE000211 Escherichia coli K-1	7.00	88.99	6.7e+03
+	gb_ba2:AR042412	11841	! AR042412 Sequence 8 from pate	7.00	88.96	6.7e+03
+	gb_pr2:AF079762	11854	! AF079762 Streptomyces venezue	7.00	88.93	6.7e+03
+	gb_vi:SINCG	12110	! J02363 Sindbis virus (hrsp an	7.00	88.88	6.8e+03
+	gb_vi:SINOCK82	12249	! M69205 Ockelbo virus complete	7.00	88.69	6.9e+03
+	gb_ba1:AE000093	12368	! AE000093 Rhizobium sp. NGR234	7.00	88.60	7.0e+03
+	gb_ba2:AE000751	12441	! AE000751 Aquifex aeolicus sec	7.00	88.59	7.0e+03
+	gb_htg5:AC017929	12457	! AC017929 Drosophila melanogas	7.00	88.58	7.1e+03
+	gb_in2:AF015193	12842	! AC017642 Onchocerca volvulus	7.00	88.47	7.1e+03
+	gb_ba2:AE000757	12959	! AE000757 Aquifex aeolicus sec	7.00	88.42	7.2e+03
+	gb_ba2:MXU24657	14003	! U24657 Myxococcus xanthus saf	7.00	88.18	7.4e+03
+	gb_htg4:AC014750	14609	! AC014750 Drosophila melanogas	7.00	88.18	7.4e+03

Right column

match	pos	score	%	7.00	strand	gb code
! AE001817 Thermotoga mariti	14621	7.4e+03	88.18	7.00	+	gb_ba2:AE001817
! AL031225 Streptomyces coel	14634	7.4e+03	88.17	7.00	+	gb_ba1:SC8B7
! AE001213 Treponema pallidu	14838	7.6e+03	88.08	7.00		gb_ba2:AE001213
! AE000340 Escherichia coli	15119	7.6e+03	87.93	7.00		gb_ba1:AE000340
! D90801 E.coli genomic DNA,	15312	7.7e+03	87.87	7.00		gb_bal:D90801
! AC010500 Homo sapiens chro	15396	7.7e+03	87.83	7.00		gb_htg7:AC010500
! AC014953 Drosophila melano	15647	7.8e+03	87.72	7.00		gb_htg4:AC014953
! D90884 E.coli genomic DNA,	15673	7.9e+03	87.71	7.00		gb_bal:D90884
! AE000709 Aquifex aeolicus	15737	7.9e+03	87.68	7.00		gb_ba2:AE000709
! AP001061 Homo sapiens geno	15894	7.9e+03	87.64	7.00		gb_prl:AP001061
! AR049868 Sequence 1 from p	15894	8.0e+03	87.62	7.00		gb_pat:AR049868
! E04903 DNA coding measles	15894	8.0e+03	87.62	7.00		gb_pat:E04903
! I59432 Sequence 1 from pat	15894	8.0e+03	87.62	7.00		gb_pat:I59432
! K01711 Measles virus (stra	15894	8.0e+03	87.62	7.00		gb_vi:MEANPCG
! I59432 Measles virus (stra	15894	8.0e+03	87.62	7.00		gb_vi:MVEDMBCDN
! S58435 nucleocapsid protei	15894	8.0e+03	87.46	7.00		gb_vi:S58435
! D90747 Escherichia coli ge	16255	8.1e+03	87.38	7.00		gb_bal:D90747
! M86514 Rat proline-rich pr	16453	8.2e+03	87.32	7.00		gb_ro:RATPRPP
! AF090337 Aythya americana	16616	8.3e+03	87.30	7.00		gb_ov:AF090337
! AR020969 Sequence 1 from p	16656	8.3e+03	87.30	7.00		gb_pat:AR020969
! AR043384 Sequence 1 from p	16656	8.3e+03	87.30	7.00		gb_pat:AR043384
! AR062299 Sequence 1 from p	16656	8.3e+03	87.11	7.00		gb_pat:AR062299
! D90885 E.coli genomic DNA,	17134	8.5e+03	87.11	7.00		gb_bal:D90885
! U00037 Caenorhabditis eleg	17172	8.5e+03	87.08	7.00		gb_in1:CELT20H4
! D90746 Escherichia coli ge	17212	8.5e+03	87.08	7.00		gb_bal:D90746
! D90754 Escherichia coli ge	17470	8.6e+03	86.98	7.00		gb_bal:D90754
! AC019546 Drosophila melano	17834	8.8e+03	86.84	7.00		gb_htg6:AC019546
! Z14056 P.solanacearum DNA	18299	8.9e+03	86.67	7.00		em_ba2:PSHRPB
! Z98600 S.pombe chromosome	18437	9.0e+03	86.62	7.00		gb_pl1:SPAC20G4
! Z27081 Caenorhabditis eleg	19001	9.3e+03	86.42	7.00		gb_in1:CEMO1A8
! Z70720 S.pombe chromosome	19108	9.3e+03	86.38	7.00		gb_pl1:SPAC1B9
! U56902 Citrus tristeza vir	19226	9.4e+03	86.34	7.00		gb_vi:CT56902
! D90802 E.coli genomic DNA,	19291	9.4e+03	86.32	7.00		gb_bal:D90802
! AF121000 Corynebacterium g	19751	9.6e+03	86.16	7.00		gb_ba2:AF121000
! AC014863 Drosophila melano	20225	9.8e+03	86.00	7.00		gb_htg4:AC014863
! X72086 Variola minor virus	22059	9.8e+03	85.99	7.00		gb_vi:VMVGEN
! Z95120 Mycobacterium tuber	22070	1.1e+04	85.41	7.00		gb_ba1:MTCY7D11
! AC019941 Drosophila melano	22411	1.1e+04	85.31	7.00		gb_htg6:AC019941
! AC019823 Drosophila melano	22500	1.1e+04	85.28	7.00		gb_htg6:AC019823
! AC023376 Drosophila melano	22822	1.1e+04	85.19	7.00		gb_htg12:AC023376
! Z68007 Caenorhabditis eleg	22887	1.1e+04	85.17	7.00		gb_in1:CERO1H5
! D16561 Human gene for ATP	23065	1.1e+04	85.12	7.00		gb_pr1:HUMATPSG
! Y92859 L.japonicus gln1, P	23326	1.1e+04	85.04	7.00		gb_pl1:LJGLN1
! AJ245811 Ralstonia solanac	23407	1.1e+04	85.02	7.00		gb_bal:RSO245811
! Y11187 A.thaliana RH1, TC1	23918	1.2e+04	84.87	7.00		gb_pl1:AT23KB
! D83992 Fission yeast DNA f	24752	1.2e+04	84.64	7.00		gb_pl1:YSP72HA
! U00049 Caenorhabditis eleg	24936	1.2e+04	84.59	7.00		gb_in2:CELC18F10
! AC015331 Drosophila melano	25014	1.2e+04	84.57	7.00		gb_htg4:AC015331
! AC006810 Caenorhabditis eleg	25201	1.2e+04	84.52	7.00		gb_htg3:AC006810
! AC008484 Homo sapiens chro	25207	1.2e+04	84.52	7.00		gb_htg7:AC008484
! X67117 Variola virus (XhoI	25344	1.2e+04	84.48	7.00		gb_vi:VVXHOIFOH
! AL133424 Streptomyces coel	25362	1.2e+04	84.48	7.00		gb_bal:SCF6
! AL032638 Caenorhabditis eleg	25377	1.2e+04	84.48	7.00		gb_in1:CEY52B11B
! AF169829 Mus musculus glyc	25662	1.3e+04	84.40	7.00		gb_ro:AF169829
! AC022098 Homo sapiens chro	25664	1.3e+04	84.40	7.00		gb_pr2:AC022098
! AF016413 Caenorhabditis el	25700	1.3e+04	84.39	7.00		gb_in1:CELF27E11
! AR068044 Sequence 5 from p	26700	1.3e+04	84.13	7.00		gb_pat:AR068044
! I91962 Sequence 1 from pat	26700	1.3e+04	84.13	7.00		gb_pat:I91962
! U79702 Mycobacterium tuber	27230	1.3e+04	84.00	7.00		gb_ba1:MTCY3G12
! U64862 Caenorhabditis eleg	27551	1.3e+04	83.92	7.00		gb_in1:CELZC8
! Z69638 Human DNA sequence	27551	1.3e+04	83.82	7.00		gb_pr2:HSL174G8A
! Z46936 Human cosmid cG1160	27974	1.3e+04	83.76	7.00		gb_pr2:HSCG1160
! Z68222 Caenorhabditis eleg	28230	1.3e+04	83.60	7.00		gb_in1:CE8ZK1251
! Z92826 Caenorhabditis eleg	28904	1.3e+04	83.52	7.00		gb_in1:CEC18D11
! AL021428 Mycobacterium tub	29249	1.3e+04	83.52	7.00		gb_bal:MTV030
! U40941 Caenorhabditis eleg	29256	1.4e+04	83.35	7.00		gb_in1:CELF35C8
! AL138538 Streptomyces coel	29992	1.4e+04	83.32	7.00		gb_ba2:SC6D10
! AC017487 Drosophila melano	30150	1.4e+04	83.16	7.00		gb_htg5:AC017487
! U18341 Variola virus Somal	30867	1.4e+04	83.16	7.00		gb_in1:VVU18341
! Z69713 Human DNA sequence	30875	1.4e+04	83.16	7.00		gb_pr3:HSN20A6
! Z71266 Caenorhabditis eleg	30911	1.4e+04	83.15	7.00		gb_in1:CERO6C7
! J02714 Chicken embryonic m	31111	1.4e+04	83.11	7.00		gb_ov:CHKMYHE
! AL021481 Caenorhabditis el	31244	1.4e+04	83.08	7.00		gb_in1:CEY43F4B

Left panel:

Name	+/-	(min)	Score	E	Position	Description
gb_in1:CELC39E6	+	7.00	83.07	1.4e+04	31280	U49944 Caenorhabditis elegans
gb_in2:CELF07E5	-	7.00	83.07	1.4e+04	31310	U80837 Caenorhabditis elegans
gb_htg4:AC012876	+	7.00	83.02	1.4e+04	31527	AC012876 Drosophila melanogas
gb_in1:VVU18339	+	7.00	83.01	1.4e+04	31562	U18339 Variola virus Garcia-1
gb_in1:CEF54G8	+	7.00	83.00	1.4e+04	31613	Z19155 Caenorhabditis elegans
gb_htg3:AC006663	-	7.00	82.88	1.5e+04	32209	AC006663 Caenorhabditis elega
gb_pr3:AF012825	+	7.00	82.85	1.5e+04	32218	AF012825 Ectromelia virus env
gb_pr2:HS306A4	+	7.00	82.83	1.5e+04	32446	AL008727 Human DNA sequence f
gb_in1:CEF40F12	-	7.00	82.74	1.5e+04	32880	Z30215 Caenorhabditis elegans
gb_bal:MTY15C10	+	7.00	82.70	1.5e+04	33050	Z95436 Mycobacterium tubercul
gb_htg6:AC018478	-	7.00	82.64	1.5e+04	33358	AC018478 Drosophila melanogas
gb_htg4:AC014874	+	7.00	82.54	1.5e+04	33855	AC014874 Drosophila melanogas
gb_ro:MSYNDE1A	-	7.00	82.53	1.5e+04	33934	Z22532 M.musculus syndecan-1.
gb_pr2:CNS00YVB	-	7.00	82.48	1.5e+04	34175	AL096805 Homo sapiens genomic
gb_htg7:AC016552	+	7.00	82.41	1.6e+04	34530	AC016552 Homo sapiens chromos
gb_in1:CEW09C5	-	7.00	82.34	1.6e+04	34873	Z82077 Caenorhabditis elegans
gb_in2:CELC16A3	+	7.00	82.32	1.6e+04	34968	U41534 Caenorhabditis elegans
gb_in2:CELZC53	-	7.00	82.31	1.6e+04	35061	U47144 Caenorhabditis elegans
gb_pr3:HS1740M1	+	7.00	82.25	1.6e+04	35369	AL035361 Homo sapiens *** SEC
gb_in1:CELF56A11	+	7.00	82.25	1.6e+04	35382	AF038619 Caenorhabditis elega
gb_pl2:SPCC188	-	7.00	82.24	1.6e+04	35412	AL049662 S.pombe chromosome I
gb_bal:SCF43A	+	7.00	82.23	1.6e+04	35437	AL096837 Streptomyces coelic
gb_in1:LMFP1105	-	7.00	82.23	1.6e+04	35443	AL139794 Leishmania major Fri
gb_bal:MLCB2407	-	7.00	82.20	1.6e+04	35615	AL023596 Mycobacterium leprae
gb_htg4:AC014835	+	7.00	82.20	1.6e+04	35541	AC014835 Drosophila melanogas
gb_pl2:SPBC16H5	+	7.00	82.19	1.6e+04	35660	AL022104 S.pombe chromosome I
gb_in1:CE2C15	+	7.00	82.15	1.6e+04	35875	Z93396 Caenorhabditis elegans
gb_in1:CELF38B6	+	7.00	82.12	1.6e+04	36044	U40060 Caenorhabditis elegans
gb_in2:CELT27A1	-	7.00	82.11	1.6e+04	36075	AF025468 Caenorhabditis elega
gb_in2:CELW03F9	+	7.00	82.11	1.6e+04	36079	AF022981 Caenorhabditis elega
gb_htg10:AC023154	-	7.00	82.11	1.6e+04	36102	AC023316 Caenorhabditis elega
gb_pl2:SPAC14C4	+	7.00	82.10	1.6e+04	36138	Z98596 S.pombe chromosome I c
gb_bal:U00014		7.00	82.04	1.6e+04	36470	U00014 Mycobacterium leprae c
gb_in1:CELW03G9	+	7.00	82.02	1.6e+04	36571	AF039710 Caenorhabditis elega
gb_pl2:AF090447	-	7.00	82.02	1.6e+04	36590	AF090447 Zea mays cosmid II.2
gb_in1:AC004646	+	7.00	81.99	1.6e+04	36767	AC004646 Homo sapiens chromos
gb_pr1:AP001062	-	7.00	81.96	1.6e+04	36931	AP001062 Homo sapiens genomic
gb_pr1:AP000543	-	7.00	81.95	1.6e+04	36955	AP000543 Homo sapiens genomic
gb_pr3:AC004700	+	7.00	81.93	1.7e+04	37056	AC004700 Homo sapiens chromos
gb_in1:AC004494	+	7.00	81.83	1.7e+04	37648	AC004494 Homo sapiens chromos
gb_htg3:AC011553	+	7.00	81.78	1.7e+04	37918	AC011553 Homo sapiens chromos
gb_pl2:SPAC10F6	+	7.00	81.74	1.7e+04	38139	AL009197 S.pombe chromosome I
gb_htg4:AC015118	-	7.00	81.74	1.7e+04	38162	AC015118 Drosophila melanogas
gb_htg3:AC010946	+	7.00	81.72	1.7e+04	38274	AC010946 Homo sapiens chromos
gb_pr1:AP000265	+	7.00	81.71	1.7e+04	38304	AP000265 Homo sapiens genomic
gb_in1:CEC56A3	-	7.00	81.64	1.7e+04	38744	Z77655 Caenorhabditis elegans
gb_bal:MSGB937CS	+	7.00	81.61	1.7e+04	38914	L78820 Mycobacterium leprae c
gb_pr4:AC005947	+	7.00	81.61	1.7e+04	38919	AC005947 Homo sapiens chromos
gb_pr4:AC006147	-	7.00	81.59	1.7e+04	38991	AC006147 Homo sapiens PAC clo
gb_pr1:AP000544	+	7.00	81.59	1.7e+04	38993	AP000544 Homo sapiens genomic
gb_htg3:AC008301	-	7.00	81.57	1.7e+04	39118	AC008301 Drosophila melanogas
gb_bal:MTCY21B4	+	7.00	81.57	1.7e+04	39150	Z80108 Mycobacterium tubercul
gb_pr3:AC002107	+	7.00	81.55	1.7e+04	39150	AC002107 Genomic sequence fro
gb_pl1:SC9395	-	7.00	81.53	1.7e+04	39265	Z46727 S.cerevisiae chromosom
gb_bal:MTCY164	+	7.00	81.50	1.7e+04	39334	Z95150 Mycobacterium tubercul
gb_pr4:AF024534	-	7.00	81.48	1.7e+04	39531	AF024534 Homo sapiens cosmid
gb_in2:CELF53A9	+	7.00	81.47	1.8e+04	39626	U23523 Caenorhabditis elegans
gb_in2:CELC23H3	-	7.00	81.43	1.8e+04	39721	U80453 Caenorhabditis elegans
gb_htg3:AC015568	+	7.00	81.41	1.8e+04	39970	AC015568 Caenorhabditis elega
gb_in2:CELC00G5	+	7.00	81.41	1.8e+04	40050	AF098985 Caenorhabditis elega
gb_htg10:AC023107	-	7.00	81.39	1.8e+04	40188	AC023107 Homo sapiens chromos
gb_bal:SCI8	+	7.00	81.35	1.8e+04	40442	AL132644 Streptomyces coelic
gb_pr3:AC004679	+	7.00	81.28	1.8e+04	40881	AC004679 Homo sapiens chromos
gb_in1:CELB0432	-	7.00	81.27	1.8e+04	40921	U80836 Caenorhabditis elegans
gb_htg3:AC006729	-	7.00	81.23	1.8e+04	41172	AC006729 Caenorhabditis elega
gb_pr1:AP000523	+	7.00	81.21	1.8e+04	41306	AP000523 Homo sapiens genomic
gb_htg5:AC017942	+	7.00	81.20	1.8e+04	41338	AC017942 Drosophila melanogas
gb_in2:CELF52F10	+	7.00	81.18	1.8e+04	41490	AF101315 Caenorhabditis elega
gb_pr3:AC004224	-	7.00	81.18	1.8e+04	41495	AC004224 Homo sapiens chromos
gb_pr3:AC004017	+	7.00	81.16	1.8e+04	41601	AC004017 Homo sapiens chromos
gb_in1:LMFL5341	-	7.00	81.15	1.8e+04	41651	AL138518 Leishmania major Fri
gb_vi:VACRHF	+	7.00	81.08	1.8e+04	42090	D11079 Vaccinia virus genomic
gb_pr3:AC001644	-	7.00	81.03	1.8e+04	42415	AC001644 Genomic sequence fro

Right panel:

Name	Description	Position	E	Score	(min)	+/-
gb_pl2:AF010283	AF010283 Sorghum bicolor A	42446	1.8e+04	81.02	7.00	+
gb_in2:CELC37C3	U64857 Caenorhabditis eleg	42521	1.8e+04	81.01	7.00	-
gb_pr3:AC005615	AC005615 Homo sapiens chro	42619	1.9e+04	80.92	7.00	+
gb_vi:HEHCMVU	280902 Human DNA sequence	43084	1.9e+04	80.89	7.00	-
gb_vi:HSN80H12	X04650 Human cytomegalovir	43275	1.9e+04	80.88	7.00	-
gb_htg4:AC014886	AC014886 Drosophila melano	43345	1.9e+04	80.88	7.00	+
gb_ov:AF083221	AF083221 Fugu rubripes put	43373	1.9e+04	80.83	7.00	+
gb_htg3:AC004396	AC004396 Homo sapiens,***	43686	1.9e+04	80.83	7.00	+
gb_pr3:AC003980	AC003980 Human Cosmid g512	43797	1.9e+04	80.81	7.00	+
gb_htg2:AC006106	AC006106 Homo sapiens chro	43798	1.9e+04	80.81	7.00	-
gb_in1:CELB0563	U28740 Caenorhabditis eleg	43801	1.9e+04	80.81	7.00	-
gb_htg6:AC021109	AC021109 Homo sapiens clon	44162	1.9e+04	80.76	7.00	+
gb_in1:CER07B1	Z48621 Caenorhabditis eleg	44209	1.9e+04	80.75	7.00	+
gb_pr2:HS426N21	Z82208 Human DNA sequence	44826	1.9e+04	80.66	7.00	-
gb_htg4:AC013883	AC013883 Drosophila melano	45371	2.0e+04	80.58	7.00	+
gb_htg8:AC020901	AC020901 Homo sapiens clon	45803	2.0e+04	80.51	7.00	-
gb_htg3:AC006610	AC006610 Caenorhabditis el	45867	2.0e+04	80.50	7.00	-
gb_pr3:AC005387	AC005387 Homo sapiens chro	46213	2.0e+04	80.45	7.00	+
gb_htg7:AC021858	AC021858 Homo sapiens clon	46215	2.0e+04	80.45	7.00	+
gb_ov:AF082667	AF082667 Gallus gallus cla	46304	2.0e+04	80.44	7.00	-
gb_pr4:AF165141	AF165141 Homo sapiens chro	46901	2.0e+04	80.35	7.00	-
gb_htg4:AC013581	AC013581 Homo sapiens clon	47064	2.0e+04	80.33	7.00	+
gb_pl3:AC007119	AC007119 Arabidopsis thali	47724	2.0e+04	80.24	7.00	+
gb_htg6:AC020726	AC020726 Homo sapiens clon	47726	2.0e+04	80.24	7.00	-
gb_htg5:AC011555	AC011555 Homo sapiens chro	48487	2.1e+04	80.13	7.00	-
gb_vi:CVI15035	Y15035 Cowpox virus strain	49649	2.1e+04	79.97	7.00	-
gb_htg2:AC002414	AC002414 Homo sapiens, ***	50873	2.2e+04	79.81	7.00	-
gb_htg7:AC021887	AC021887 Homo sapiens clon	51851	2.2e+04	79.68	7.00	+
gb_pr2:HS27C10	AL031803 Human DNA sequenc	52358	2.2e+04	79.62	7.00	+
gb_htg2:AC003656_7	Continuation (8 of 8) of A	53000	2.2e+04	79.53	7.00	-
gb_in2:CELT12A2	U13019 Caenorhabditis eleg	54118	2.3e+04	79.39	7.00	+
gb_htg12:AC023868	AC023868 Homo sapiens chro	55178	2.3e+04	79.26	7.00	-
gb_pl2:SCD9509	U32274 Saccharomyces cerev	55359	2.3e+04	79.24	7.00	-
gb_htg11:AC012761	AC012761 Drosophila melano	55551	2.3e+04	79.22	7.00	-
gb_htg11:AC023375	AC023375 Homo sapiens chro	56043	2.3e+04	79.16	7.00	-
gb_htg11:AC017592	AC017592 Drosophila melano	56371	2.4e+04	79.12	7.00	-
gb_htg1:HS317C6	Z97651 Homo sapiens chromo	58191	2.4e+04	78.91	7.00	+
gb_htg4:AC011722	AC011722 Homo sapiens chro	59146	2.5e+04	78.80	7.00	+
gb_htg11:AC023300	AC023300 Homo sapiens chro	59308	2.5e+04	78.78	7.00	-
gb_htg11:AC023028	AC023028 Homo sapiens chro	59725	2.5e+04	78.73	7.00	-
gb_htg11:AC023297	AC023297 Homo sapiens clon	60074	2.5e+04	78.69	7.00	-
gb_pr2:HSJ964N17	AL050311 Human DNA sequenc	60904	2.5e+04	78.60	7.00	-
gb_htg5:AC018533	AC018533 Homo sapiens chro	60917	2.5e+04	78.60	7.00	-
gb_htg7:AC009112	AC009112 Homo sapiens chro	61001	2.5e+04	78.59	7.00	-
gb_htg7:AC010235	AC010235 Homo sapiens chro	61849	2.6e+04	78.50	7.00	-
gb_htg10:AC023525	AC023525 Homo sapiens chro	63227	2.6e+04	78.35	7.00	-
gb_htg4:AC014845	AC014845 Drosophila melano	63315	2.6e+04	78.34	7.00	-
gb_htg4:AC022915	AC022915 Homo sapiens chro	63437	2.6e+04	78.33	7.00	-
gb_htg6:AC019923	AC019923 Drosophila melano	63971	2.6e+04	78.27	7.00	-
gb_in2:AC001653	AC001653 Drosophila melano	64167	2.6e+04	78.25	7.00	-
gb_bal:MTV043	AL022004 Mycobacterium tub	64687	2.7e+04	78.20	7.00	-
gb_htg10:AC022818	AC022818 Homo sapiens chro	65307	2.7e+04	78.13	7.00	-
gb_pl3:AC007155	AC007155 Arabidopsis thal	65443	2.7e+04	78.12	7.00	-
gb_htg4:AC015858	AC015858 Homo sapiens chro	66432	2.7e+04	78.02	7.00	-
gb_pr1:AB017652	AB017652 Homo sapiens geno	66497	2.7e+04	77.97	7.00	-
gb_pr1:AP000464	AP000464 Drosophila melano	66898	2.7e+04	77.96	7.00	-
gb_in2:AC004301	AC004301 Drosophila melano	66991	2.7e+04	77.92	7.00	-
gb_htg11:AC023572	AC023572 Mycobacterium tub	67435	2.8e+04	77.89	7.00	-
gb_bal:MTV043	AL022004 Mycobacterium tub	67701	2.8e+04	77.83	7.00	-
gb_htg6:AC019785	AC019785 Drosophila melano	68351	2.8e+04	77.80	7.00	-
gb_ov:AC009592	AC009592 Drosophila melano	68582	2.8e+04	77.80	7.00	+
gb_pl1:AB015469	AB015469 Arabidopsis thali	68620	2.8e+04	77.80	7.00	+
gb_htg11:AC023666	AC023666 Homo sapiens chro	68742	2.8e+04	77.78	7.00	-
gb_htg6:AC020671	AC020671 Homo sapiens chro	68848	2.8e+04	77.78	7.00	+
gb_pr2:HS4I15	AL049549 Human DNA sequenc	69541	2.8e+04	77.71	7.00	-
gb_htg3:AC007722	AC007722 Homo sapiens chro	70935	2.9e+04	77.58	7.00	-
gb_htg7:AC021422	AC021422 Homo sapiens chro	72027	2.9e+04	77.47	7.00	-

ID		7.00	Temp	Conc	Num		Accession / Description
gb_ba2:ECOUW67_3	-	7.00	77.44	2.9e+04	72438	!	Continuation (4 of 4) of ECOU
gb_htg4:AC016354	+	7.00	77.39	2.9e+04	72975	!	AC016354 Homo sapiens chromos
gb_pr2:HSDJ723E3	-	7.00	77.37	2.9e+04	73112	!	AL049736 Human DNA sequence f
gb_htg14:AC021066	+	7.00	77.37	2.9e+04	73112	!	AC021066 Homo sapiens chromos
gb_htg12:AC010563	-	7.00	77.34	2.9e+04	73461	!	AC010563 Drosophila melanogas
gb_htg10:AC022983	+	7.00	77.33	3.0e+04	73643	!	AC022983 Homo sapiens chromos
gb_htg6:AC020007	-	7.00	77.31	3.0e+04	73721	!	AC020007 Drosophila melanogas
gb_htg12:AC020405	-	7.00	77.31	3.0e+04	73834	!	AC020405 Drosophila melanogas
gb_htg6:AC020128	-	7.00	77.29	3.0e+04	74054	!	AC020128 Drosophila melanogas
gb_htg12:AC023361	+	7.00	77.22	3.0e+04	74673	!	AC023361 Mus musculus clone R
gb_pr4:AC004855	-	7.00	77.21	3.0e+04	74816	!	AC004855 Homo sapiens PAC clo
gb_htg5:AC016260	-	7.00	77.18	3.0e+04	74957	!	AC016260 Homo sapiens clone R
gb_in2:AC004114	-	7.00	77.16	3.0e+04	75307	!	AC004114 Drosophila melanogas
gb_htg4:AC011953	-	7.00	77.15	3.0e+04	75542	!	AC011953 Homo sapiens clone 1
gb_htg10:AC022903	-	7.00	77.10	3.0e+04	75558	!	AC022903 Homo sapiens chromos
gb_htg4:AC012976	-	7.00	77.09	3.0e+04	76179	!	AC012976 Drosophila melanogas
gb_htg3:AC006496	-	7.00	77.02	3.1e+04	76269	!	AC006496 Drosophila melanogas
gb_htg5:AC016879	-	7.00	77.01	3.1e+04	77071	!	AC016879 Homo sapiens chromos
gb_in2:AC004581	-	7.00	76.97	3.1e+04	77226	!	AC004581 Homo sapiens, *** SE
gb_htg11:AC023223	-	7.00	76.96	3.1e+04	77642	!	AC023223 Homo sapiens chromos
gb_in2:AC004369	-	7.00	76.96	3.1e+04	77798	!	AC004369 Drosophila melanogas
gb_in2:AC004369	-	7.00	76.94	3.1e+04	77798	!	AC004369 Drosophila melanogas
gb_in2:AC004296	-	7.00	76.92	3.1e+04	78054	!	AC004296 Drosophila melanogas
gb_pl1:AB011477	-	7.00	76.87	3.1e+04	78181	!	AB011477 Arabidopsis thaliana
gb_htg10:AC022788	-	7.00	76.85	3.1e+04	78803	!	AC022788 Homo sapiens chromos
gb_pl1:AB008267	-	7.00	76.82	3.1e+04	79046	!	AB008267 Arabidopsis thaliana
gb_htg7:AC021683	-	7.00	76.79	3.2e+04	79408	!	AC021683 Homo sapiens clone F
gb_htg4:AC015940	-	7.00	76.79	3.2e+04	79745	!	AC015940 Homo sapiens chromos
gb_htg5:AC017034	-	7.00	76.75	3.2e+04	79761	!	AC017034 Homo sapiens clone R
gb_htg7:AC021826	-	7.00	76.73	3.2e+04	80230	!	AC021826 Homo sapiens chromos
gb_pl2:ATT27B3	-	7.00	76.73	3.2e+04	80450	!	AL137079 Arabidopsis thaliana
gb_pl2:ATT9A21	-	7.00	76.73	3.2e+04	80498	!	AC021365 Homo sapiens clone R
gb_htg7:AC021365	-	7.00	76.71	3.2e+04	80537	!	AC012183 Homo sapiens chromos
gb_htg4:AC012183	-	7.00	76.70	3.2e+04	80755	!	AC004009 Homo sapiens PAC clc
gb_pr4:AC004009	-	7.00	76.66	3.2e+04	80866	!	AC012649 Drosophila melanogas
gb_htg4:AC012649	-	7.00	76.65	3.2e+04	81370	!	AB017071 Arabidopsis thaliana
gb_pl1:AB017071	-	7.00	76.64	3.2e+04	81494	!	AB019236 Arabidopsis thaliana
gb_htg9:AB019236	-	7.00	76.64	3.2e+04	81534	!	AC021275 Homo sapiens clone F
gb_htg7:AC021275	-	7.00	76.64	3.2e+04	81546	!	AC021686 Homo sapiens clone R
gb_htg10:AC021686	-	7.00	76.60	3.2e+04	81552	!	AC004359 Drosophila melanogas
gb_in1:AC004359	-	7.00	76.58	3.2e+04	82109	!	AC003925 Drosophila melanogas
gb_htg3:AC003925	-	7.00	76.57	3.3e+04	82263	!	AC007999 Homo sapiens 3q25-26
gb_pr5:AC007999	-	7.00	76.56	3.3e+04	82469	!	AC005430 Homo sapiens clone N
gb_htg4:AC005430	-	7.00	76.55	3.3e+04	82532	!	AF186999 Homo sapiens chromos
gb_htg3:AF186999	-	7.00	76.54	3.3e+04	82697	!	AL021713 Arabidopsis thaliana
gb_pl2:ATT9A21	-	7.00	76.53	3.3e+04	82814	!	AL021713 Arabidopsis thaliana
gb_pl2:ATT9A21	-	7.00	76.47	3.3e+04	83607	!	AC006836 Arabidopsis thaliana
gb_htg13:AC006836	-	7.00	76.47	3.3e+04	83650	!	AF024533 Arabidopsis thaliana
gb_pr4:AF024533	-	7.00	76.45	3.3e+04	83946	!	AC006264 Arabidopsis thaliana
gb_htg13:AC006264	-	7.00	76.40	3.3e+04	84508	!	AB015474 Genomic sequence for
gb_htg5:AB015474	-	7.00	76.40	3.3e+04	84510	!	AF027868 Bacillus subtilis ch
gb_ba2:AF027868	-	7.00	76.37	3.3e+04	84621	!	Continuation (4 of 4) of CEY3
gb_pl11:CEY37H9_3	-	7.00	76.18	3.4e+04	84912	!	AB011479 Arabidopsis thaliana
gb_pr3:HSA12M24	-	7.00	76.17	3.4e+04	87328	!	AC009955 Homo sapiens clone N
gb_htg4:AC009955	-	7.00	76.15	3.4e+04	87487	!	AB006698 Arabidopsis thaliana
gb_in1:CEY11G7B	-	7.00	76.08	3.5e+04	87500	!	AC004537 Homo sapiens PAC clc
gb_pr5:AC004537	+	7.00	76.06	3.5e+04	87805	!	AC003670 Homo sapiens 12q13.1
gb_pr3:HSJ5230	+	7.00	76.06	3.5e+04	88356	!	AL049555 Homo sapiens chromos
gb_htg2:HSJ523K23	+	7.00	76.03	3.5e+04	88662	!	AC008848 Homo sapiens chromos
gb_htg5:AC008848	+	7.00	75.97	3.5e+04	88872	!	AC005956 Arabidopsis thaliana
gb_pl3:AC005956	-	7.00	75.96	3.5e+04	88945	!	AL033523 Human DNA sequence f
gb_pr3:H72E17	-	7.00	75.94	3.5e+04	89319	!	AL008722 Human DNA sequence f
gb_htg2:AL139801	+	7.00	75.90	3.5e+04	89359	!	AL139801 Homo sapiens chromos
gb_htg7:HS26H23	-	7.00	75.88	3.5e+04	90113	!	AC008508 Homo sapiens cosmid
gb_pl1:ATF24A6	+	7.00	75.84	3.6e+04	90299	!	Z84467 Human DNA sequence frc
gb_pr3:HS390B3	+	7.00	75.83	3.6e+04	90497	!	AL035396 Arabidopsis thaliana
			75.80	3.6e+04	91028	!	Z93096 Human DNA sequence frc

	Accession / Description	Num	Conc	Temp	7.00		ID
!	AL136360 Homo sapiens chro	93097	3.6e+04	75.75	7.00	+	gb_htg1:AL136360
!	AC007269 Homo sapiens clon	93207	3.6e+04	75.74	7.00	-	gb_pr4:AC007269
!	U95737 Human Chromosome 16	93431	3.6e+04	75.73	7.00	+	gb_pr5:HUU95737
!	AC012462 Homo sapiens clon	93634	3.6e+04	75.71	7.00	+	gb_pr4:AC012462
!	AL035539 Arabidopsis thali	93760	3.6e+04	75.70	7.00	-	gb_pl2:ATF22I13
!	AC010386 Homo sapiens chro	93780	3.6e+04	75.68	7.00	-	gb_htg10:AC010386
!	AC005667 Homo sapiens clon	94106	3.7e+04	75.65	7.00	-	gb_pr3:AC005667
!	AC023347 Homo sapiens chro	94448	3.7e+04	75.64	7.00	-	gb_htg11:AC023347
!	AL096841 Homo sapiens clon	94715	3.7e+04	75.61	7.00	-	gb_htg2:HSJ577L21
!	AC008720 Homo sapiens chro	95156	3.7e+04	75.58	7.00	-	gb_htg7:AC008720
!	AP000650 Homo sapiens chro	95568	3.7e+04	75.57	7.00	-	gb_htg1:AP000650
!	AC004100 Homo sapiens chro	95604	3.7e+04	75.54	7.00	-	gb_pr3:AC004100
!	AC014559 Drosophila melano	95934	3.7e+04	75.55	7.00	+	gb_htg4:AC014559
!	AC011173 Homo sapiens geno	96076	3.7e+04	75.54	7.00	+	gb_htg4:AC011173
!	AP000243 Homo sapiens geno	96174	3.7e+04	75.53	7.00	-	gb_pr1:AP000243
!	AC011230 Homo sapiens clon	96201	3.7e+04	75.53	7.00	-	gb_htg3:AC011230
!	Z99291 Human DNA sequence	96608	3.7e+04	75.50	7.00	+	gb_pr2:HS308I13
!	AL121813 Drosophila melano	96924	3.7e+04	75.48	7.00	+	gb_htg1:DMBR2123
!	AC004906 Homo sapiens clon	96964	3.7e+04	75.48	7.00	-	gb_htg2:AC004906
!	AL138645 Arabidopsis thali	96988	3.7e+04	75.48	7.00	-	gb_pl2:ATF26B15
!	AC011991 Homo sapiens clon	97075	3.7e+04	75.47	7.00	-	gb_htg4:AC011991
!	AC008633 Homo sapiens chro	97118	3.7e+04	75.47	7.00	-	gb_htg10:AC008633
!	Z97629 Human DNA sequence	98902	3.8e+04	75.35	7.00	-	gb_pr2:HS128N22
!	AC008075 Arabidopsis thali	98942	3.8e+04	75.34	7.00	-	gb_pl2:F24J5
!	AC023117 Homo sapiens gen	98954	3.8e+04	75.34	7.00	-	gb_htg10:AC023117
!	AC005773 Homo sapiens gen	99036	3.8e+04	75.34	7.00	+	gb_pr4:AC005773
!	AL139177 Homo sapiens gen	99263	3.8e+04	75.32	7.00	+	gb_htg2:CNS01DX9
!	AL035523 Arabidopsis thali	99688	3.8e+04	75.29	7.00	-	gb_pl2:ATF13M23
!	AC022352 Oryza sativa gen	99790	3.8e+04	75.29	7.00	-	gb_htg8:AC022352
!	AC005990 Arabidopsis thali	99923	3.8e+04	75.28	7.00	-	gb_pl2:F508
!	AP000013 Homo sapiens gen	100000	3.8e+04	75.27	7.00	-	gb_pr1:AP000013
!	AP000034 Homo sapiens gen	100000	3.8e+04	75.27	7.00	-	gb_pr1:AP000034
!	AP000101 Homo sapiens gen	100000	3.8e+04	75.27	7.00	-	gb_pr1:AP000101
!	AP000154 Homo sapiens gen	100000	3.8e+04	75.27	7.00	-	gb_pr1:AP000154
!	AP000177 Arabidopsis thal	100000	3.8e+04	75.27	7.00	-	gb_pr1:AP000177
!	AP000203 Homo sapiens gen	100640	3.8e+04	75.23	7.00	-	gb_pr1:AP000203
!	AC010965 Mus musculus clo	100722	3.9e+04	75.22	7.00	-	gb_htg3:AC010965
!	AC009283 Homo sapiens chr	100771	3.9e+04	75.21	7.00	-	gb_htg3:AC009283
!	AC008711 Homo sapiens chr	100960	3.9e+04	75.21	7.00	-	gb_htg7:AC008711
!	AC007980 Arabidopsis thal	100975	3.9e+04	75.21	7.00	-	gb_pl2:AC007980
!	AC018341 Drosophila melan	101085	3.9e+04	75.20	7.00	-	gb_htg5:AC018341
!	AC007905 Homo sapiens chr	100722	3.9e+04	75.18	7.00	-	gb_htg6:AC020380
!	AC011478 Homo sapiens chr	101484	3.9e+04	75.17	7.00	-	gb_htg6:AC011478
!	Z84478 Human DNA sequence	101877	3.9e+04	75.15	7.00	-	gb_pr3:HS37M3
!	AC005004 Homo sapiens PAC	103606	3.9e+04	75.03	7.00	-	gb_pr4:AC005004
!	AL136445 Homo sapiens chr	103964	4.0e+04	75.01	7.00	-	gb_htg2:AL136445
!	AC022578 Drosophila melan	104597	4.0e+04	74.97	7.00	-	gb_pr2:HS393P12
!	AC017784 Drosophila melan	104695	4.0e+04	74.96	7.00	-	gb_htg5:AC017784
!	AC004075 Homo sapiens Chr	104961	4.0e+04	74.95	7.00	+	gb_pr3:AC004075
!	AC021869 Homo sapiens clo	105868	4.0e+04	74.89	7.00	-	gb_htg5:AC021869
!	AC008486 Homo sapiens clo	105914	4.0e+04	74.89	7.00	-	gb_htg7:AC020380
!	AC011478 Homo sapiens chr	106018	4.0e+04	74.88	7.00	-	gb_pr2:HS355N11
!	AL009051 Homo sapiens DNA	106319	4.0e+04	74.86	7.00	-	gb_pr3:HS747L4
!	AC015814 Homo sapiens clo	106362	4.0e+04	74.86	7.00	-	gb_pr4:AC015814
!	AC017423 Drosophila melan	106551	4.0e+04	74.85	7.00	-	gb_htg5:AC017423
!	AC011436 Arabidopsis thal	106688	4.1e+04	74.84	7.00	-	gb_pl2:ATAC011436
!	AC007691 Drosophila melan	106926	4.1e+04	74.82	7.00	-	gb_htg3:AC007691
!	AL133232 Homo sapiens chr	107266	4.1e+04	74.80	7.00	-	gb_htg2:AL133232
!	AC022383 Homo sapiens chr	108141	4.1e+04	74.75	7.00	-	gb_htg9:AC022383
!	Continuation (4 of 4) of	108350	4.1e+04	74.73	7.00	-	gb_htg1:HSG118G19_3
!	D10879 Herpes simplex vir	108395	4.1e+04	74.73	7.00	+	gb_v1:HS1ULR
!	AC020591 Homo sapiens clo	108488	4.1e+04	74.72	7.00	-	gb_htg6:AC020591
!	AC000159 Homo sapiens Chr	108589	4.1e+04	74.72	7.00	-	gb_htg3:AC000159
!	AC006795 Caenorhabditis e	109106	4.1e+04	74.69	7.00	-	gb_htg3:AC006795
!	AC000674 Homo sapiens chr	109210	4.1e+04	74.68	7.00	-	gb_htg1:AC000674
!	AL049761 Human BAC clone	109389	4.1e+04	74.67	7.00	-	gb_htg7:HSDJ863C7
!	AC002530 Human DNA sequen	109430	4.1e+04	74.67	7.00	-	gb_pr5:AC002530
!	AC012434 Homo sapiens clo	109488	4.1e+04	74.66	7.00	-	gb_htg11:AC012434
!	AL031774 Human DNA sequen	109575	4.1e+04	74.66	7.00	-	gb_pr3:HSJ1069P2
!	AL109839 Human DNA sequen	109758	4.1e+04	74.65	7.00	-	gb_htg12:AC022222
!	AJ246003 Homo sapiens Spa	110000	4.2e+04	74.63	7.00	+	gb_pr2:HSA246003

This page contains a dense two-block tabular data listing (a sequence alignment / oligo report). Each row has the form: sequence identifier, strand (+/−), numeric score columns, a continuation or descriptive text field, a length value, and an expectation value (e.g. 4.2e+04).

Left block (selected rows, columns: ID | strand | 7.00 | value | 7.00 | description | length | e-value):

ID	±		value		description	length	e-value
gb_htg1:CEY113G7_2	−	7.00	74.63	7.00	! Continuation (3 of 8) of CEY	110000	4.2e+04
gb_htg1:CEY113G7_3	+	7.00	74.63	7.00	! Continuation (4 of 8) of CEY	110000	4.2e+04
gb_htg1:CEY53H1_0	+	7.00	74.63	7.00	! Z99775 Caenorhabditis elegan	110000	4.2e+04
gb_htg1:HSG118G19_0	+	7.00	74.63	7.00	! AL121996 Homo sapiens chromo	110000	4.2e+04
gb_htg1:HSMX1_5	−	7.00	74.63	7.00	! Continuation (6 of 7) of HSM	110000	4.2e+04

Right block (selected rows, columns: gb_description | strand | 7.00 | value | e-value | length | accession | organism text):

gb_description	±		value	e-value	length	accession	organism
gb_pr3:HS503G16	−	7.00	73.74	4.7e+04	125686	Z93020	Human DNA sequence
gb_pl1:ATF2K18	+	7.00	73.73	4.7e+04	125803	AL035556	Arabidopsis thal
gb_htg1:AP000666	+	7.00	73.72	4.7e+04	125983	AP000666	Homo sapiens chr
gb_htg7:AC019032	+	7.00	73.71	4.7e+04	126189	AC019032	Homo sapiens chr
gb_htg4:AC009392	+	7.00	73.71	4.7e+04	126227	AC009392	Drosophila melan

(The page continues with approximately 80 similar rows of sequence identifiers, scores, accession numbers such as AC004934, AC004139, AC007243, AC010679, and organism descriptions including "Homo sapiens chr", "Drosophila melanoga", "Caenorhabditis e", "Mus musculus", "Synechocystis sp.", and "Human DNA sequen". Numeric score columns range from approximately 73.11–73.74 and 74.63, with e-values of 4.2e+04–5.0e+04 on the right block and 4.2e+04–4.6e+04 on the left block.)

```
gb_pr3:HSJ876B10    +  7.00  73.11  5.0e+04  ! AL117352 Human DNA sequence
gb_pr4:AC004908     +  7.00  73.10  5.0e+04  ! AC004908 Homo sapiens PAC cl
gb_pr4:AC005280     +  7.00  73.09  5.1e+04  ! AC005280 Homo sapiens clone
gb_ht2:AC002979     +  7.00  73.09  5.1e+04  ! AC002979 Homo sapiens, *** S
gb_htg1:HSJ539E24   +  7.00  73.08  5.1e+04  ! AL121920 Homo sapiens chromo
gb_htg11:AC012388   +  7.00  73.07  5.1e+04  ! AC012388 Drosophila melanoga
gb_htg10:AC010668   -  7.00  73.06  5.1e+04  ! AC010668 Drosophila melanoga
gb_htg2:AL137062    -  7.00  73.05  5.1e+04  ! AL137062 Homo sapiens chromo
gb_pr3:HS68O2       +  7.00  73.04  5.1e+04  ! Z82215 Human DNA sequence fr
gb_ba1:D90905       -  7.00  73.04  5.1e+04  ! D90905 Synechocystis sp. PCC
gb_htg3:AC008025    +  7.00  73.04  5.1e+04  ! AC008025 Homo sapiens clone
gb_htg7:AC019315    +  7.00  73.03  5.1e+04  ! AC019315 Homo sapiens chromo
```

seq_name: gb_ba2:AE000213

```
seq_documentation_block:
LOCUS       AE000213   10959 bp    DNA           BCT       12-NOV-1998
DEFINITION  Escherichia coli K-12 MG1655 section 103 of 400 of the complete
            genome.
ACCESSION   AE000213 U00096
VERSION     AE000213.1  GI:1787371
KEYWORDS    .
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 10959)
AUTHORS     Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
TITLE       The complete genome sequence of Escherichia coli K-12
JOURNAL     Science 277 (5331), 1453-1474 (1997)
MEDLINE     97426617
REFERENCE   2  (bases 1 to 10959)
AUTHORS     Blattner,F.R.
TITLE       Direct Submission
JOURNAL     Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 10959)
AUTHORS     Blattner,F.R.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 10959)
AUTHORS     Plunkett,G. III.
TITLE       Direct Submission
JOURNAL     Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K-12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 (e-mail: mark@amber.gatech.edu). Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site.Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K-12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
```

```
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1..10959
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:562"
     promoter        13..40
                     /note="factor Sigma70; predicted +1 start at 1184959"
     protein_bind    119..136
                     /note="central position to predicted promoter:82"
                     /bound_moiety="PhoB predicted site"
     gene            155..1381
                     /gene="pepT"
                     /note="b1127"
     CDS             155..1381
                     /gene="pepT"
                     /EC_number="3.4.11.-"
                     /function="putative enzyme; Not classified"
                     /note="o408; 100 pct identical to 42 aa fragment
                     PEPT_ECOLISW: P29745; 93 pct identical to 407 residues of
                     409 aa PEPT_SALTY SW: P26311"
                     /codon_start=1
                     /transl_table=11
                     /product="putative peptidase T"
                     /protein_id="AAC74211.1"
                     /db_xref="GI:1787372"
                     /db_xref="PID:g1787372"
                     /translation="MDKLLERFLNYVSLDTQSKAGVRQVPSTEQQWKLLHLLKEQLEE
                     MGLINVTLSEKGTLMATLPANVPGDIPAIGFISHVDTSPDCSGKNVNPQIVENYRGGD
                     IALGIGDEVLSPVMFPVLHQLLGQTLITTDGKTLGADDKAGIAEIMTALAVLQQKKI
                     PHGDIRVAFTPDEEVGKGAKHFDVDAFDARWAYTVDGGGVGELEFENFNAASVNKIV
                     GNNVHPGTAKGVMVNALSLAARIHAEVPADESPEMTEGYEGFYHLASMKGTVERADMH
                     YIIRDFDRKQFEARKRKMMEIAKKVGKGLHPDCYIELVIEDSYYNMREKVEHPHILD
                     IAQQAMRDCDIEPELKPIRGGTDGAQLSMGLPCPNLFTGGYNYHGKHEFVTLEGMEK
                     AVQVIVRAELTAQRK"
     gene            complement(1430..2560)
                     /gene="ycfD"
                     /note="b1128"
     CDS             complement(1430..2560)
                     /gene="ycfD"
                     /function="orf; Unknown"
                     /note="f376; 100 pct identical to fragment YCFD_ECOLI
                     SW:P27431 but has 3 additional N-terminal aa and 93
                     additional C-terminal aa"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74212.1"
                     /db_xref="GI:1787373"
                     /db_xref="PID:g1787373"
                     /translation="MLNMEYQLTLNWPDFLERHWQKRPVVLKRGFNNFIDPISPDELA
                     GLAMESEVDSRLVSHQDGKWQVSHGPFESYDHLGETNWSLLVQAVNHWHEPTAALMRP
                     FRELPDWRIDDLMISFSVPGGVGPHLDQYDVFIIQGTGRRRWRVGEKLQMKQHCPHP
                     DLLQVDPFEAIIDELEPGDILYIPPGFPHEGYALENAMNYSVGFRAPNTRELISGFA
                     DYVLQRELGGNYYSDPDVPPRAHPADVLPQEMDKLREMMLELINQPEHFKQWFGEFIS
                     QSRHELDIAPPEPPYQPDEIYDALKQGEVLVRLGGLRVLRIGDDVYANGEKIDSPHRP
                     ALDALASNIALTAENFGDALEDPSFLAMLAALVNSGYWFEG"
     promoter        complement(2583..2613)
                     /note="factor Sigma70; predicted +1 start at 1187488"
     gene            complement(2627..4087)
                     /gene="phoQ"
                     /note="b1129"
     CDS             complement(2627..4087)
                     /gene="phoQ"
                     /EC_number="2.7.3.-"
                     /function="enzyme; Global regulatory functions"
```

```
                              Ratio:  1.000                    Gaps:        0
          Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AE000213/rev

Align seg 1/1 to reverse of: AE000213  from: 1 to: 10959

   151  TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProThr   167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
  8818  TGGTTATGGCCGGAATCACCAATTCGTGAACGCAAAAGTATTCCCAC         8769

   167  rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA     184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
  8768  CAGCTGGCTGGAAGATCACCTTATATGAAGACGTAATCGCCAGGTGCGCC      8719

   184  rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla    200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
  8718  GCATGACCGCCCATGTGGCTTCCCACCGCTGACGACTGAATCGCTATGCG     8669

   201  MetGly  202
        ||||||
  8668  ATGGGT  8663

seq_name: gb_bal:D90748

seq_documentation_block:
LOCUS       D90748      15007 bp   DNA                         BCT     07-FEB-1999
DEFINITION  Escherichia coli genomic DNA. (25.6 - 25.9 min).
ACCESSION   D90748 AB001340
VERSION     D90748.1 GI:1651553
KEYWORDS    Complete and shotgun sequencing; potB; potA; pepT; phoQ; phoP;
            purB; YcfC; ycfB; lcdA; lcd; lcdE; lit.
SOURCE      Escherichia coli(strain:K12) DNA. clone:Kohara clone #239.
ORGANISM    Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 15007)
AUTHORS     Mori,H.
TITLE       Direct Submission
JOURNAL     Submitted (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
            Hirotada Mori, NARA Institute of Science and Technology, Res. 6
            Edu. Center for Genetic Info.: 8916-5 Takayama, Ikoma, Nara 630-01,
            Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
REFERENCE   2 (sites)
AUTHORS     Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito.N., Sampei.G., Seki.Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
TITLE       The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL     Unpublished (1996)
REFERENCE   3 (sites)
AUTHORS     Oshima.T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito.N., Sampei.G., Seki.Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
TITLE       A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL     DNA Res. 3 (3), 137-155 (1996)
MEDLINE     97061202
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
```

```
              /note="f486; 99 pct identical to PHOQ_ECOLI SW: P23837"
              /codon_start=1
              /transl_table=11
              /product="sensor protein PhoQ"
              /protein_id="AAC74213.1"
              /db_xref="GI:1787374"
              /db_xref="PID:g1787374"
              /translation="MKKLLRLFFPLSLRVRFLLATAAVVLVLSLAYGMVALIGYSVSF
              DKTTFRLLRGESNLFYLAKWENNKLHVELPENIDKQSPTMTLIYDENGQLLWAQRDV
              PWLMKMIQPDWLKSNGFHEIEADVNDTSLLLSGDHSIQQQLQEVREDDDDAEMTHSVA
              VNVYPATSRMPKLTIVVDTIPVELKSSYMVWSWFIYVLSANLLLVIPLLWVAAWWSL
              RPIEALAKEVRELEEHNRELLNPATTRELTSLVRNLNRLLKSERERYDKYRTLTDLT
              HSLKTPLAVLQSTLRSLRSEKMSVSDAEPVMLEQISRISQIGYLHRASMRGGTLLS
              RELHPVAPLLDNLTSALNKVYQRKGVNISLDISPEISFVGEQNDFVEVMGNVLDNACK
              ICLEFVEISARQTDEHLYIVVEDDGPGIPLSKREVIFDRGQRVDTLRPGQGVGLAVAR
              EITEQYRGKIVAGESMLGGARMEVIFGRQHSAPKDE"
  gene        complement(4087..4758)
              /gene="phoP"
              /note="b1130"
  CDS         complement(4087..4758)
              /gene="phoP"
              /function="regulator; Global regulatory functions"
              /note="f223; 100 pct identical to PHOP_ECOLI SW: P23836"
              /codon_start=1
              /transl_table=11
              /product="transcriptional regulatory protein"
              /protein_id="AAC74214.1"
              /db_xref="GI:1787375"
              /db_xref="PID:g1787375"
              /translation="MRVLVVEDNALLRHHLKVQIQDAGHQVDDAEDAKEADYLNEHI
              PDIAIVDLGLPDEDGLSLIRRWRSNDVSLPIVLTARESWQDKVEVLSAGADDYVTKP
              FHIEEVVARMQALMRRNSGLASQVISLPFFQVDLSRRELSINDEVIKLTAFEYTIMET
              LIRNNGKVVSKDSLMLQLYPDAELRESHTIDVLMGRLRKKIQAQYPQEVITTVRGQGY
              LFELR"
  promoter    complement(4257..4287)
              /gene="phoP"
              /note="factor sigma70; predicted +1 start at 1189162"
  gene        complement(4927..6297)
              /gene="purB"
              /note="b1131"
  CDS         complement(4927..6297)
              /gene="purB"
              /EC_number="4.3.2.2"
              /function="enzyme; Purine ribonucleotide biosynthesis"
              /note="f456; 99 pct identical to PURB_ECOLI SW: P25739"
              /codon_start=1
              /transl_table=11
              /product="adenylosuccinate lyase"
              /protein_id="AAC74215.1"
              /db_xref="GI:1787376"
              /db_xref="PID:g1787376"
              /translation="MELSSLTAVSPVDGRYGDKVSALRGIFSEYGLLKFRVQVEVRWL
              QKLAAHAAIKEVPAFAADAIGYLDAIVASFSEEDAARIKTIERTTNHDVKAVEYFLKE
              KVAEIPELHAVSEFIHFACTSEDINNLSHALMKTARDEVILPYWRQLIDGIKDLAVQ
              YRDIPLLSRTHGQPATPSTIGKEMANVAYRMERQYRQLNQVEILGKINGAVGNYNAHI
              AAYPEVDWHQFSEEFVTSLGIOWNPYTTQIEPHDYIAELFDCVARFNTILIDFDRDVW
              GYIALNFKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNAVLQHLASKLPVSRWQ
              RDLTDSTVLRNLGVGIGYALIAYQSTLKGVSKLEVNRDHLLDELDHNWEVLAEPIQTV
              MRRYGIEKPYEKLKELTRGKRVDAEGMKQFIDGLALPEEEKARLKAMTPANYIGRAIT
              MVDELK"
  gene        complement(6301..6942)
              /gene="ycfC"
              /note="b1132"
  CDS         complement(6301..6942)
              /gene="ycfC"
              /function="orf; Unknown"
              /note="f213; 99 pct identical to YCFC_ECOLI SW: P25746"
              /codon_start=1
              /transl_table=11
              /product="orf; hypothetical protein"
```

```
alignment_scores:
  Quality:  52.00        Length:        52
```

Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotada Mori
Address: NARA Institute of Science and Technology,
         Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.

FEATURES   Location/Qualifiers
source   1..15007
                      /organism="Escherichia coli"
                      /strain="K12"
                      /db_xref="taxon:562"
                      /clone="Kohara clone #239"
                      /map="25.6-25.9 min"
                      /note="Nucleotide position 1185468-1200474 from the
                      initiation site of ThrA (0 min.).; This clone is from
                      Kohara lambda miniset library."
gene   complement(1..678)
                      /gene="potB"
CDS   complement(<1..678)
                      /gene="potB"
                      /note="ORF_ID:o238#14; similar to SwissProt Accession
                      Number P23860"
                      /codon_start=1
                      /transl_table=11
                      /product="Spermidine/putrescine transport system permease
                      protein PotB."
                      /protein_id="BAA35947.1"
                      /db_xref="GI:1651554"
                      /translation="MIVTIVGWLVLFVFLPNLMIIGTSFLTRDDASFVMVFTLDNYT
                      RLLDDLPFEVLLHSLNMALIATLACLVLGYPFAWFLAKLPHKVRPLLLFLLIVPFWTN
                      SLIRIYGLKIFLSTKGYLNEFLLWLGVIDTPIRIMFTPSAVIIGLVYILLPFMVMPLY
                      SSIEKLDKPLEAARDLGASKLQTFIRIIIPLTMPGIIAGCLLVMLPAMGLFYVSDLM
                      GGAKNLLI"
gene   complement(692..1828)
                      /gene="potA"
CDS   complement(692..1828)
                      /gene="potA"
                      /note="ORF_ID:o239#1; similar to PIR Accession Number
                      A40840"
                      /codon_start=1
                      /transl_table=11
                      /product="Spermidine/putrescine transport protein A"
                      /protein_id="BAA35948.1"
                      /db_xref="GI:1651555"
                      /translation="MGQSKKLNKQPSSLSPLVQLAGIRKCFDGKEVIPQLDLTINNGE
                      FLTLLGPSGCGKTTVLRLIAGLETVDSGRIMLDNEDITHVPAENRYVNTVFQSYALFP
                      HMTVFENVAFGLRMQKTPAAEITPRVMEALRMVQLETFAQRKPHQLSGGQQQRVAIAR
                      AVVNKPRLLLLDESLSALDYKLRKQMNELKALQRKLGITFVFVTHDQEALTMSDRI
                      VVMRDGRIEQDGTPREIYEPKNLFVAGFIGEINMFNATVIERLDEQRVRANVEGREC
                      NGKMVMVSEFNEDDPDFDHSLDQKMAINWVESWEVVLADEEHR"
gene   2078..3304
                      /gene="pepT"
CDS   2078..3304
                      /gene="pepT"
                      /note="ORF_ID:o239#2; similar to SwissProt Accession
                      Number P29745"
                      /codon_start=1
                      /transl_table=11
                      /product="Peptidase T (EC 3.4.11.-) (aminotripeptidase)

(tripeptidase)."
                      /protein_id="BAA35949.1"
                      /db_xref="GI:1651556"
                      /translation="MDKLLERFLNVVSLDTQSKAGVRQVPSTEGQWKLLHLKEQLEE
                      MGLINVTLSEKGTLMATLPANVPGDIPAIGFISHVDTSPDCSGKNVNPQIVENYRGGD
                      IALGIGDEVLSPVMPPVLHQLLGQTLITTDGKTLLGADDKAGIAEIMTALAVLQQKKI
                      PHGDIRVAFTPDEEVGKGAKHFDVDAFDARWAYTVDGGVGELEFENFNAASVNIKIV
                      GNNVHPGTAKGVMVNALSLAARIHAEVPADESPEMTEGYEGFYHLASMKGTVERADMH
                      YIIRDFDRKQFEARKRKMMEIAKKVGKGLHPDCYIELVIEDSYYNMREKVVEHPILD
                      IAQQAMRDCDIEPELKPIRGGTDGAQLSFWGLPCPNLFTGGYNYHGKHEFVTLEGMEK
                      AVQVIVRIAELTAQRK"
gene   complement(3353..6010)
                      /gene="phoQ"
CDS   complement(3353..4474)
                      /gene="phoQ"
                      /note="ORF_ID:o239#3; similar to PIR Accession Number
                      D41966"
                      /codon_start=1
                      /transl_table=11
                      /product="Orf 2 downstream of phoQ."
                      /protein_id="BAA35950.1"
                      /db_xref="GI:4062695"
                      /translation="MEYQLTLNWPDFLERHWQKRPVVLKRGFNNFIDPISPDELAGLA
                      MESEVDSRLVSHQDGKWQVSHGPPESYDHLGETNWSLLVQAVNHWHEPTAALMRPFRE
                      LPDWRIDDLMISFSVPGGVGVPHLDQYDVFIIQGTGRRRWRVGERKLQMKQHCPHPDLL
                      QVDPPEAIIDEELEPGDILYIPPGFPHEGVALENAMNYSVGFRAPNTRELISGFADYV
                      LQRELGGNYSDPDVPPRAHPADVLPQEMDKLREMMLELINQPEHFKQWFGEFISQSR
                      HELDIAPPEPYQPDEIYDALKQGEVLVRLGGLRVLRIGDDVYANGEKIDSPHRPALD
                      ALASNIALTAENFGDALEDDSFLAMLAALVNSGYWFFEG"
gene   complement(4550..6010)
                      /gene="phoQ"
CDS   /note="ORF_ID:o239#4; similar to PIR Accession Number
                      B41966"
                      /codon_start=1
                      /transl_table=11
                      /product="Virulence membrane protein phoQ."
                      /protein_id="BAA35951.1"
                      /db_xref="GI:1651557"
                      /translation="MKKLRLFFPLSLRVRFLLATAAVVLVLSLAYGMVALIGYSVSF
                      DKTFRLLRGESNLFYTLAKWENNKLHVELPENIDKQSPTMTLIYDENGQLLWAQRDV
                      PWLMKMIQPDWLKSNGFHEIEADVNDTSLLSGDHSIQQOLQEVFEDDDAEMTHSVA
                      VNVTPATSRMPKLTIVVVDTIPVELKSSYMVWSWFIYVLSANLLLVIPLLWVAAWWSL
                      RPIRALAKEVRELEEHNRELNLNPATTRELTSLVRNLNRLLKSERERYDKRTTLTDLT
                      HSLKTPLAVLQSTLRSLRSEKMSVSDAEPVMLEQISRISQIGYLHRASMRGGTLLS
                      RELHPVAPLLDNLTSALNKVYQRGVNISLDISPEISFVGEQNDFVEVMGNVLDNACK
                      YCLEFVEISARQTDEHLYIVVEDDGPGIPLSKREVIFDRGQRVDTLRPGGQGVGLAVAR
                      EITEYEGKIVAGESMLGGARMEVLFGRQHSAPKDE"
gene   complement(6010..6681)
                      /gene="phoP"
CDS   complement(6010..6681)
                      /gene="phoP"
                      /note="ORF_ID:o240#1; similar to SwissProt Accession
                      Number P23836"
                      /codon_start=1
                      /transl_table=11
                      /product="Transcriptional regulatory protein PhoP."
                      /protein_id="BAA35952.1"
                      /db_xref="GI:1651558"
                      /translation="MRVLVVEDNALLRHHLKVQIQDAGHQVDDAEDAKEADYLNEHI
                      PDIAVDLGLPDEDGLSLIRRWRSNDVSLPILVLTARESWQDKVEVLSAGADDYVTKP
                      FHIESVMARMQALMRRNSGLASQVISLPPFQVDLSRRELSINDEVIKLTAFEYIMET
                      LIRNNGKVVSKDSLMLQLYPDAELRESHTIDVLMGRLRRKIQAQYPEVTTVRGQGY
                      LFELR"
gene   complement(6850..8220)
                      /gene="purB"
CDS   complement(6850..8220)
                      /gene="purB"
                      /note="ORF_ID:o240#2; similar to PIR Accession Number
                      S19212"
                      /codon_start=1
                      /transl_table=11
                      /product="Adenylosuccinate lyase (EC 4.3.2.2)"
                      /protein_id="BAA35953.1"

/db_xref="GI:1651559"
/translation="MELSSLTAVSPVDGRYGDKVSALRGIFSEYGLLKPRVQVEVRWL
QKLAHAAIKEVPAFAADAIGYLDAIVASFSEEDAARIKTIERTNHDVKAVEYFLKE
KVAEIPELHAVSEFTHFACTSEDINNLSHALMLKTARDEVILPYWRQLIDGIKDLAVQ
YRDIPLLSRTHGQPATPSTIGKEMANVAYRMERQYRQLNQVEILGKINGAVGNYNAHI
AAYPEVDWHQFSEEFVTSLGIQWNPYTTQEPHDYIAELFDCVARFNTILIDFDRDVW
GYIALNHFKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNAVLQHLASKLPVSRWQ
RDLTDSTVLRNLGVGIGYALIAYQSTLKGVSKLEVNRDHLLDELDHNWEVLAEPIQTV

alignment_scores:
    Quality: 52.00   Length: 52
    Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x D90748/rev   ..

Align seg 1/1 to reverse of: D90748 from: 1 to: 15007

    151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProTh 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  10741 TGGTTATGGCCGCGGAATCCACCAATTCGTGAACGCAAAAGTATTCCCAC 10692

    167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  10691 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGCC 10642

    184 rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  10641 GCATGACCGCCCATGTTGGCTTCCCCACGCTGCGACTGATTCGCTATGCG 10592

    201 MetGly 202
        ||||||
  10591 ATGGGT 10586

seq_name: gb_bal:D90749

seq_documentation_block:
LOCUS   D90749   20284 bp   DNA   BCT   07-FEB-1999
DEFINITION Escherichia coli genomic DNA. (25.7 - 26.1 min).
ACCESSION   D90749 AB001340
VERSION   D90749.1 GI:1651562
KEYWORDS   Complete and shotgun sequencing; phoQ; phoP; purB; ycfC; ycfB;
            icdA; icd; icdE; lit; int; xis; ycfK; ycfE; pin; ycfA.
SOURCE   Escherichia coli(strain:K12) DNA, clone:Kohara clone #240.
ORGANISM   Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 20284)
AUTHORS   Mori,H.
TITLE   Direct Submission
JOURNAL   Submitted (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
            Hirotada Mori, NARA Institute of Science and Technology, Res. &
            Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01,
            Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
REFERENCE   2 (sites)
AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kimura,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
TITLE   The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL   Unpublished (1996)
REFERENCE   3 (sites)
AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.

TITLE   A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL   DNA Res. 3 (3), 137-155 (1996)
MEDLINE   97061202
COMMENT   Collaboration Information:
            Project:
            The Japan E.coli genome DNA sequencing project
            Group:
            The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
            Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
            Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
            Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
            Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
            Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
            Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
            Headed by:
            Name: Takashi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kishori@nibb.ac.jp
            Information operator:
            Name: Hirotada Mori
            Address: NARA Institute of Science and Technology,
            Ikoma, 630-01, Japan
            E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
            The Japan E. coli genome database
            http:bsw3.aist-nara.ac.jp.
            Location/Qualifiers
FEATURES
     source   1..20284
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /clone="Kohara clone #240"
                     /map="25.7-26.1 min"
                     /note="Nucleotide position 1191151-1211434 from the
                     initiation site of ThrA (0 min.).; This clone is from
                     Kohara lambda miniset library."
     gene   complement(1. .327)
                     /gene="phoQ"
     CDS   complement(<1. .327)
                     /gene="phoQ"
                     /note="ORF_ID:o239#4; similar to PIR Accession Number
                     B41966"
                     /codon_start=1
                     /transl_table=11
                     /product="Virulence membrane protein phoQ."
                     /protein_id="BAA35960.1"
                     /db_xref="GI:1651563"
                     /translation="MKKLLRLFFPLSLRVRFLLATAAVVLVLSLAYGMVALIGYSVSF
                     DKTTFRLRGESNLFYTLAKWENNKLHVELPENIDKQSPTMTLIYDENGQLLWAQRDV
                     PWLMKMI"
     gene   complement(327. .998)
                     /gene="phoP"
     CDS   complement(327. .998)
                     /gene="phoP"
                     /note="ORF_ID:o240#1; similar to SwissProt Accession
                     Number P23836"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulatory protein PhoP."
                     /protein_id="BAA35961.1"
                     /db_xref="GI:1651564"
                     /translation="MRVLVVEDNALLRHHLKVQIQDAGHQVDDAEDAKEADYLNEHI
                     PDIAIVDLGLPDEDGLSLIRRWRSNDVSLPIIVLTARESWQDKVEVLSAGADDYVTKP
                     FHIEEVMARMQALMRRNSGLASQVISLPPFQVDLSRRELSINDEVIKTAFEYTIMET
                     LIRNNGKVVSKDSLMLQLYPDAELRESHTIDVLMGRLRKKIQAQYPQEVITTVRGQGY
                     LFELR"
     gene   complement(1167. .2537)
                     /gene="purB"
     CDS   complement(1167. .2537)

```
gene        KSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMGDYSLDNLANGEWREVTD"
            5574..6924
            /gene="icd"
CDS         5574..6924
            /gene="icd"
            /note="ORF_ID:o240#7; similar to SwissProt Accession
            Number P08200"
            /codon_start=1
            /transl_table=11
            /product="isocitrate dehydrogenase (NADP) (EC  1.1.1.42)
            (oxalosuccinate decarboxylase) (idh) (NADP+-specific icdh)
            (idp)."
            /protein_id="BAA35967.1"
            /db_xref="GI:1651566"
            /translation="MESKVVVPAQGKKITLQNGKLNVPENPIIPYIEGDGIGVDVTPA
            MLKVVDAAVEKAYKGERKISWMEIYTGEKSTQVYGQDVWLPAETLDLIREYRVAIKGP
            LTTPVGGGIRSLNVALRQELDYICLRPVRYQGTPSPVKHPELTDMVIFRNSEDIY
            AGIEWKADSADAEKVIKFLREEMGVKKIRFPEHCGIGIKPCSEEGTKRLVRAAIEYAI
            ANDRDSVTLVHKGNIMKFTEGAFKDWGYQLAREEFGGELIDGGPWLKVKNPNTGKEIV
```

```
alignment_scores:
        Quality:       52.00        Length:        52
        Ratio:         1.000        Gaps:          0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x D90749/rev   ..
Align seg 1/1   to reverse of: D90749  from: 1  to: 20284

     151 TrpLeuThrProArgAsnProProIleArgGluArgLysSerIleProTh 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    5058 TGGTTATGCCGCGGGGAATCCACCAATTCGTGAACGCAAAAGTATTCCCAC 5009

     167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    5008 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGCC 4959

     184 rgMetThrAlaHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    4958 GCATGACCGCCCATGTTGGCTTCCCCACGCTGCGACTGATTCGCTATGCG 4909

     201 MetGlyTyr 202
         |||||||
    4908 ATGGGT 4903
```

```
seq_name: gb_htg7:AC022161

seq_documentation_block:
LOCUS       AC022161     180993 bp     DNA              HTG        26-JAN-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-273P11, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC022161
VERSION     AC022161.1  GI:6758614
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180993)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180993)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     ------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            ------
```

```
            /gene="purB"
            /note="ORF_ID:o240#2; similar to PIR Accession Number
            S19212"
            /codon_start=1
            /transl_table=11
            /product="Adenylosuccinate lyase (EC  4.3.2.2)"
            /protein_id="BAA35962.1"
            /db_xref="GI:1651565"
            /translation="MELSSLTAVSPVDGRYGDKVSALRGIFSEYGLLKFRVQVEVRWL
            QKLAAHAIKEVPAFADAIGTLDATVASFSEEDAARIKTIERTTNHDVRAVEYFLKE
            KVAEIPELHAVSEFIHFACTSEDINNLSHALMLKTARDEVILPYWRQLIDGIKDLAVQ
            YRDIPLLSRTHGQPATPSTIGKEMANVAYRMERQYRQLNQVEILGKINGAVGNYNAHI
            AAYPEVDWHQFSEEFVTSLGIQWNPYTTQIEPHDYIAELFDCVARFNTILIDFDRDVW
            GYIALNHFKQKTIAGEIGSSTMPHKVNPIDFENSEGNLGLSNAVLQHLASKLPVSRWQ
            RDLTDSTVLRNLGVGIGYALIAYQSTLKGVSKLEVNRDHLLDELDRNWEVLAEPIQTV
            MRRYGIEKPYEKLKELTRGKRVDAEGMKQFIDGLALPEEEKARLKAMTPANYIGRAIT
            MVDELK"
gene        complement(2541..3182)
            /gene="ycfC"
CDS         complement(2541..3182)
            /gene="ycfC"
            /note="ORF_ID:o240#3; similar to PIR Accession Number
            S19211"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein 23."
            /protein_id="BAA35963.1"
            /db_xref="GI:4062714"
            /translation="MAKNYDITLALAGICQSARLVQQLAHQGHCDADALHVSLNSII
            DMNPSSTLAVFGGSEANLRVGLETLLGVLNASSRQGLNAELTRYTISLMVLERKLSSA
            KGALDTLGNRINGLQRQLEHFDLQSETLMSAMAAIVDVISPLGPRIQVTGSPAVLQS
            POVQAKVRATLLAGIRAAVLMHQVGGGRLQLMFSRNRLTTQAKQILAHLTPEL"
gene        complement(3218..4324)
            /gene="ycfB"
CDS         complement(3218..4324)
            /gene="ycfB"
            /note="ORF_ID:o240#4; similar to SwissProt Accession
            Number P25745"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein in purB 5'region (orf-15)
            "
            /protein_id="BAA35964.1"
            /db_xref="GI:4062715"
            /translation="MSETAKKVIVGMSGGVDSSVSAWLLQQGIQVEGLFMKNWEEDD
            GEEYCTAAADLADAQAVCKDLKIGIELHTVNFAAEYWDNVFELFLAEYKAGRTPNPDLLC
            NKEIKPKAFLEFAAEDLGADYIATGHYVRRADVDGKSRLRGLDSNKDQSYFLYTLSH
            EQIAQSLFPVGELEKPQVRKIAEDLGLVTAKKKDSTGICFIGERKFREFLGRYLPAQP
            GKIITVDGDEIGEHQGLMYHIGQRGLGIGGTKEGTEEPWYVVDRDVENNILVVAQG
            HEHPRLMSVGLIAQQLHWVDREPFTGTMRCTVKTRYRQTDIPCTVKALDDDRIEVIFD
            EPVAAVTPGQSAVFYNGEVCLGGGIIEQRLPLPV"
CDS         complement(4378..4839)
            /note="ORF_ID:o240#5; similar to PIR Accession Number
            S64074"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein YGL067w"
            /protein_id="BAA35965.1"
            /db_xref="GI:4062716"
            /translation="MFKPHVTVACVVHAEGKFLVVEETINGKALWNQPAGHLEADETL
            VEAAARELWEETGISAQPQHFIRMHQWIAPDKTPFLRFLFFAIELEICPTQPHDSDID
            CCRWVSAEEILQASNLRSPLVAESIRCYQSGQRYPLEMIGDFNWPFTKGVI"
CDS         complement(4849..5502)
            /note="ORF_ID:o240#6; similar to PIR Accession Number
            I64156"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein HI0694"
            /protein_id="BAA35966.1"
            /db_xref="GI:4062717"
            /translation="MRQFIISENTMQKTSFRNHQVKRFSSQRSTRRKPENQPTRVILF
            NKPYDVLPQFTDEAGRKTLKEFIPVQGVYAAGRLDRDSEGLLVLTNNGALQARLTQPG
            KRTGKIYYVQVEGIPTQDALALRNGVTLNDGPTLPAGAELVDEPAWLWPRNPPIRER
```

```
* NOTE: This record contains 113 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
*       1    511: contig of 511 bp in length
*              gap of unknown length
*     512   1507: contig of 996 bp in length
*              gap of unknown length
*    1508   2279: contig of 772 bp in length
*              gap of unknown length
*    2280   2614: contig of 335 bp in length
*              gap of unknown length
*    2615   3450: contig of 836 bp in length
*              gap of unknown length
*    3451   3914: contig of 464 bp in length
*              gap of unknown length
*    3915   4858: contig of 944 bp in length
*              gap of unknown length
*    4859   5380: contig of 522 bp in length
*              gap of unknown length
*    5381   6247: contig of 867 bp in length
*              gap of unknown length
*    6248   7203: contig of 956 bp in length
*              gap of unknown length
*    7204   7920: contig of 717 bp in length
*              gap of unknown length
*    7921   8924: contig of 1004 bp in length
*              gap of unknown length
*    8925   9674: contig of 750 bp in length
*              gap of unknown length
*    9675  11049: contig of 1375 bp in length
*              gap of unknown length
*   11050  11834: contig of 785 bp in length
*              gap of unknown length
*   11835  12774: contig of 940 bp in length
*              gap of unknown length
*   12775  13555: contig of 781 bp in length
*              gap of unknown length
*   13556  14538: contig of 983 bp in length
*              gap of unknown length
*   14539  15398: contig of 860 bp in length
*              gap of unknown length
*   15399  15696: contig of 298 bp in length
*              gap of unknown length
*   15697  16675: contig of 979 bp in length
*              gap of unknown length
*   16676  17597: contig of 922 bp in length
*              gap of unknown length
*   17598  18969: contig of 1372 bp in length
*              gap of unknown length
*   18970  19300: contig of 331 bp in length
*              gap of unknown length
*   19301  19380: contig of 80 bp in length
*              gap of unknown length
*   19381  20457: contig of 1077 bp in length
*              gap of unknown length
*   20458  21187: contig of 730 bp in length
*              gap of unknown length
*   21188  21435: contig of 248 bp in length
*              gap of unknown length
*   21436  22621: contig of 1186 bp in length
*              gap of unknown length
*   22622  23209: contig of 588 bp in length
*              gap of unknown length
*   23210  23994: contig of 785 bp in length
*              gap of unknown length
*   23995  24643: contig of 649 bp in length
*              gap of unknown length
*   24644  25421: contig of 778 bp in length
*              gap of unknown length
*   25422  25826: contig of 405 bp in length
*              gap of unknown length
*   25827  27532: contig of 1706 bp in length
*              gap of unknown length
*   27533  28937: contig of 1405 bp in length
*              gap of unknown length
*   28938  30011: contig of 1074 bp in length
*              gap of unknown length
*   30012  30574: contig of 563 bp in length
*              gap of unknown length
*   30575  31545: contig of 971 bp in length
*              gap of unknown length
*   31546  32314: contig of 769 bp in length
*              gap of unknown length
*   32315  33010: contig of 696 bp in length
*              gap of unknown length
*   33011  33654: contig of 644 bp in length
*              gap of unknown length
*   33655  34705: contig of 1051 bp in length
*              gap of unknown length
*   34706  35755: contig of 1050 bp in length
*              gap of unknown length
*   35756  36672: contig of 917 bp in length
*              gap of unknown length
*   36673  37904: contig of 1232 bp in length
*              gap of unknown length
*   37905  38457: contig of 553 bp in length
*              gap of unknown length
*   38458  39301: contig of 844 bp in length
*              gap of unknown length
*   39302  40257: contig of 956 bp in length
*              gap of unknown length
*   40258  41691: contig of 1434 bp in length
*              gap of unknown length
*   41692  43227: contig of 1536 bp in length
*              gap of unknown length
*   43228  44416: contig of 1189 bp in length
*              gap of unknown length
*   44417  45206: contig of 790 bp in length
*              gap of unknown length
*   45207  46223: contig of 1017 bp in length
*              gap of unknown length
*   46224  47104: contig of 881 bp in length
*              gap of unknown length
*   47105  49150: contig of 2046 bp in length
*              gap of unknown length
*   49151  50227: contig of 1077 bp in length
*              gap of unknown length
*   50228  51990: contig of 1763 bp in length
*              gap of unknown length
*   51991  52651: contig of 661 bp in length
*              gap of unknown length
*   52652  54157: contig of 1506 bp in length
*              gap of unknown length
*   54158  55684: contig of 1527 bp in length
*              gap of unknown length
*   55685  56936: contig of 1252 bp in length
*              gap of unknown length
*   56937  57993: contig of 1057 bp in length
*              gap of unknown length
*   57994  58875: contig of 882 bp in length
*              gap of unknown length
*   58876  60147: contig of 1272 bp in length
*              gap of unknown length
*   60148  61314: contig of 1167 bp in length
*              gap of unknown length
*   61315  61502: contig of 188 bp in length
*              gap of unknown length
*   61503  62298: contig of 796 bp in length
```

JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
          On Jan 14, 2000 this sequence version replaced gi:5686162.

COMMENT   -----Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          -----Summary Statistics
          Consensus quality: 120029 bases at least Q40
          Consensus quality: 163743 bases at least Q30
          Consensus quality: 177506 bases at least Q20
          Estimated insert size: 226805; sum-of-contigs estimation
          Estimated insert size: 164170; agarose-fp estimation
          Quality coverage: 4.03x in Q20 bases; agarose-fp estimation
          Quality coverage: 2.91x in Q20 bases; sum-of-contigs estimation
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 95 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.

          *   1   1018: contig of 1018 bp in length
          *   gap of unknown length
          *   1019   2070: contig of 1052 bp in length
          *   gap of unknown length
          *   2071   3108: contig of 1038 bp in length
          *   gap of unknown length
          *   3109   4211: contig of 1103 bp in length
          *   gap of unknown length
          *   4212   5226: contig of 1015 bp in length
          *   gap of unknown length
          *   5227   6336: contig of 1110 bp in length
          *   gap of unknown length
          *   6337   7420: contig of 1084 bp in length
          *   gap of unknown length
          *   7421   8436: contig of 1016 bp in length
          *   gap of unknown length
          *   8437   9514: contig of 1078 bp in length
          *   gap of unknown length
          *   9515 11442: contig of 1928 bp in length
          *   gap of unknown length
          *   11443 12908: contig of 1466 bp in length
          *   gap of unknown length
          *   12909 14243: contig of 1335 bp in length
          *   gap of unknown length
          *   14244 15370: contig of 1127 bp in length
          *   gap of unknown length
          *   15371 16374: contig of 1004 bp in length
          *   gap of unknown length
          *   16375 17484: contig of 1110 bp in length
          *   gap of unknown length
          *   17485 18486: contig of 1002 bp in length
          *   gap of unknown length
          *   18487 19535: contig of 1049 bp in length
          *   gap of unknown length
          *   19536 20747: contig of 1212 bp in length
          *   gap of unknown length
          *   20748 21794: contig of 1047 bp in length
          *   gap of unknown length
          *   21795 22910: contig of 1116 bp in length
          *   gap of unknown length
          *   22911 24510: contig of 1600 bp in length
          *   gap of unknown length
          *   24511 26330: contig of 1820 bp in length
          *   gap of unknown length
          *   26331 27876: contig of 1546 bp in length
          *   gap of unknown length
          *   27877 29050: contig of 1174 bp in length
          *   gap of unknown length
          *   29051 30329: contig of 1279 bp in length

          *   62299 63529: contig of 1231 bp in length
          *   gap of unknown length
          *   63530 65376: contig of 1847 bp in length
          *   gap of unknown length
          *   65377 65461: contig of 85 bp in length
          *   gap of unknown length
          *   65462 66798: contig of 1337 bp in length
          *   gap of unknown length
          *   66799 68065: contig of 1267 bp in length
          *   gap of unknown length
          *   68066 69742: contig of 1677 bp in length
          *   gap of unknown length
          *   69743 71264: contig of 1522 bp in length
          *   gap of unknown length
          *   71265 72180: contig of 916 bp in length
          *   gap of unknown length
          *   72181 74061: contig of 1881 bp in length
          *   gap of unknown length
          *   74062 75047: contig of 986 bp in length
          *   gap of unknown length
          *   75048 76521: contig of 1474 bp in length
          *   gap of unknown length
          *   76522 78628: contig of 2107 bp in length
          *   gap of unknown length
          *   78629 80347: contig of 1719 bp in length
          *   gap of unknown length
          *   80348 82199: contig of 1852 bp in length
          *   gap of unknown length
          *   82200 83804: contig of 1605 bp in length

alignment_scores:
        Quality: 42.00   Length: 42
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AC022161

Align seg 1/1   to: AC022161   from: 1   to: 180993

    151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProThr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
  32188 TGGTTATGGCCGCGGAATCCACCAATTCGTGAACGCAAAGTATTCCCAC 32237

    167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
  32238 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGCC 32287

    184 rgMetThrAlaHisValGlyPhePro 192
        ||||||||||||||||||||||||||
  32288 GCATGACCGCCCCATGTTGGCTTCCCC 32313

seq_name: gb_htg6:AC008911

seq_documentation_block:
LOCUS   AC008911 226805 bp   DNA   HTG   14-JAN-2000
DEFINITION Homo sapiens chromosome 5 clone CITB-H1_226BJ5, WORKING DRAFT
            SEQUENCE, 95 unordered pieces.
ACCESSION   AC008911
VERSION   AC008911.2 GI:6693307
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 226805)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 226805)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission

us-09-252-691-7056.oligo.rge

Tue Jun 13 10:29:41 2000

```
* 30330       gap of unknown length
* 31614:  contig of 1285 bp in length
* 31615       gap of unknown length
* 33078:  contig of 1464 bp in length
* 33079       gap of unknown length
* 34288:  contig of 1210 bp in length
* 34289       gap of unknown length
* 35341:  contig of 1053 bp in length
* 35342       gap of unknown length
* 36695:  contig of 1354 bp in length
* 36696       gap of unknown length
* 37729:  contig of 1034 bp in length
* 37730       gap of unknown length
* 39224:  contig of 1495 bp in length
* 39225       gap of unknown length
* 40393:  contig of 1169 bp in length
* 40394       gap of unknown length
* 41828:  contig of 1435 bp in length
* 41829       gap of unknown length
* 43639:  contig of 1811 bp in length
* 43640       gap of unknown length
* 45023:  contig of 1384 bp in length
* 45024       gap of unknown length
* 46157:  contig of 1134 bp in length
* 46158       gap of unknown length
* 47250:  contig of 1093 bp in length
* 47251       gap of unknown length
* 48869:  contig of 1619 bp in length
* 48870       gap of unknown length
* 50138:  contig of 1269 bp in length
* 50139       gap of unknown length
* 51169:  contig of 1031 bp in length
* 51170       gap of unknown length
* 52255:  contig of 1086 bp in length
* 52256       gap of unknown length
* 53454:  contig of 1199 bp in length
* 53455       gap of unknown length
* 55052:  contig of 1598 bp in length
* 55053       gap of unknown length
* 56384:  contig of 1332 bp in length
* 56385       gap of unknown length
* 57426:  contig of 1042 bp in length
* 57427       gap of unknown length
* 58894:  contig of 1468 bp in length
* 58895       gap of unknown length
* 59899:  contig of 1005 bp in length
* 59900       gap of unknown length
* 61097:  contig of 1198 bp in length
* 61098       gap of unknown length
* 62827:  contig of 1730 bp in length
* 62828       gap of unknown length
* 64143:  contig of 1316 bp in length
* 64144       gap of unknown length
* 65320:  contig of 1177 bp in length
* 65321       gap of unknown length
* 66338:  contig of 1018 bp in length
* 66339       gap of unknown length
* 67938:  contig of 1600 bp in length
* 67939       gap of unknown length
* 69534:  contig of 1596 bp in length
* 69535       gap of unknown length
* 70638:  contig of 1104 bp in length
* 70639       gap of unknown length
* 71721:  contig of 1083 bp in length
* 71722       gap of unknown length
* 72737:  contig of 1016 bp in length
* 72738       gap of unknown length
* 74240:  contig of 1503 bp in length
* 74241       gap of unknown length
* 76005:  contig of 1765 bp in length
* 76006       gap of unknown length
* 77033:  contig of 1028 bp in length

*  77034       gap of unknown length
*  79330:  contig of 2297 bp in length
*  79331       gap of unknown length
*  80402:  contig of 1072 bp in length
*  80403       gap of unknown length
*  82066:  contig of 1664 bp in length
*  82067       gap of unknown length
*  83385:  contig of 1319 bp in length
*  83386       gap of unknown length
*  85506:  contig of 2121 bp in length
*  85507       gap of unknown length
*  87079:  contig of 1573 bp in length
*  87080       gap of unknown length
*  88558:  contig of 1479 bp in length
*  88559       gap of unknown length
*  89799:  contig of 1241 bp in length
*  89800       gap of unknown length
*  92035:  contig of 2236 bp in length
*  92036       gap of unknown length
*  94093:  contig of 2058 bp in length
*  94094       gap of unknown length
*  96588:  contig of 2495 bp in length
*  96589       gap of unknown length
*  98564:  contig of 1976 bp in length
*  98565       gap of unknown length
* 100900:  contig of 2336 bp in length
* 100901       gap of unknown length
* 103171:  contig of 2271 bp in length
* 103172       gap of unknown length
* 105268:  contig of 2097 bp in length
* 105269       gap of unknown length
* 108853:  contig of 3585 bp in length
* 108854       gap of unknown length
* 112992:  contig of 4139 bp in length
* 112993       gap of unknown length
* 117661:  contig of 4669 bp in length
* 117662       gap of unknown length
* 121578:  contig of 3917 bp in length


alignment_scores:
         Quality:  28.00                  Length:      28
           Ratio:   1.000                   Gaps:       0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AC008911/rev   ..

Align seg 1/1  to reverse of: AC008911  from: 1  to: 226805

  151 TrpLeuTrpProArgAsnProProIleArgArgLysSerIleProTh 167
      |||||||||||||||||||||||||||||||||||||||||||||||||
50738 TGGTTATGGCCGCCGGAATCCACCAATTCGTGAACGCAAAGTATTCCCAC 50689

  167 rSerTrpLeuLysIleThrLeuTyrGluGlyArg 178
      ||||||||||||||||||||||||||||||||||
50688 CAGCTGGCTGAAGATCACCCTTATATGAAGGACGC 50655

seq_name: gb_htg7:AC009127

seq_documentation_block:
LOCUS       AC009127     186591 bp    DNA          HTG       26-JAN-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-498D10, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC009127
VERSION     AC009127.2  GI:6758895
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 186591)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
```

REFERENCE   2 (bases 1 to 185591)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Jan 26, 2000 this sequence version replaced gi:5685948.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            * NOTE: This record contains 101 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *   1   989: contig of 989 bp in length
            *   990   gap of unknown length
            *   990   1318: contig of 329 bp in length
            *   1319   gap of unknown length
            *   1319   1843: contig of 525 bp in length
            *   1844   gap of unknown length
            *   1844   2066: contig of 223 bp in length
            *   2067   gap of unknown length
            *   2067   3127: contig of 1061 bp in length
            *   3128   gap of unknown length
            *   3128   3535: contig of 408 bp in length
            *   3536   gap of unknown length
            *   3536   4171: contig of 636 bp in length
            *   4172   gap of unknown length
            *   4172   4797: contig of 626 bp in length
            *   4798   gap of unknown length
            *   4798   5170: contig of 373 bp in length
            *   5171   gap of unknown length
            *   5171   5390: contig of 220 bp in length
            *   5391   gap of unknown length
            *   5391   6021: contig of 631 bp in length
            *   6022   gap of unknown length
            *   6022   6686: contig of 665 bp in length
            *   6687   gap of unknown length
            *   6687   6975: contig of 289 bp in length
            *   6976   gap of unknown length
            *   6976   7331: contig of 356 bp in length
            *   7332   gap of unknown length
            *   7332   7911: contig of 580 bp in length
            *   7912   gap of unknown length
            *   7912   8248: contig of 337 bp in length
            *   8249   gap of unknown length
            *   8249   8844: contig of 596 bp in length
            *   8845   gap of unknown length
            *   8845   9149: contig of 305 bp in length
            *   9150   gap of unknown length
            *   9150   9469: contig of 320 bp in length
            *   9470   gap of unknown length
            *   9470   9911: contig of 442 bp in length
            *   9912   gap of unknown length
            *   9912 10494: contig of 583 bp in length
            *   10495   gap of unknown length
            *   10495 11143: contig of 649 bp in length
            *   11144   gap of unknown length
            *   11144 11762: contig of 619 bp in length
            *   11763   gap of unknown length
            *   11763 12380: contig of 618 bp in length
            *   12381   gap of unknown length
            *   12381 13000: contig of 620 bp in length
            *   13001   gap of unknown length
            *   13001 13576: contig of 576 bp in length
            *   13577   gap of unknown length
            *   13577 14175: contig of 599 bp in length
            *   14176   gap of unknown length
            *   14176 14801: contig of 626 bp in length
            *   14802   gap of unknown length
            *   14802 15373: contig of 572 bp in length
            *   15374   gap of unknown length
            *   15374 15716: contig of 343 bp in length
            *   15717   gap of unknown length
            *   15717 16489: contig of 773 bp in length
            *   16490   gap of unknown length
            *   16490 17111: contig of 622 bp in length
            *   17112   gap of unknown length
            *   17112 17855: contig of 744 bp in length
            *   17856   gap of unknown length
            *   17856 19178: contig of 1323 bp in length
            *   19179   gap of unknown length
            *   19179 20052: contig of 874 bp in length
            *   20053   gap of unknown length
            *   20053 20790: contig of 738 bp in length
            *   20791   gap of unknown length
            *   20791 21464: contig of 674 bp in length
            *   21465   gap of unknown length
            *   21465 22165: contig of 701 bp in length
            *   22166   gap of unknown length
            *   22166 22966: contig of 801 bp in length
            *   22967   gap of unknown length
            *   22967 23458: contig of 492 bp in length
            *   23459   gap of unknown length
            *   23459 24138: contig of 680 bp in length
            *   24139   gap of unknown length
            *   24139 24984: contig of 846 bp in length
            *   24985   gap of unknown length
            *   24985 25642: contig of 658 bp in length
            *   25643   gap of unknown length
            *   25643 26565: contig of 923 bp in length
            *   26566   gap of unknown length
            *   26566 27196: contig of 631 bp in length
            *   27197   gap of unknown length
            *   27197 27815: contig of 619 bp in length
            *   27816   gap of unknown length
            *   27816 27901: contig of 86 bp in length
            *   27902   gap of unknown length
            *   27902 28512: contig of 611 bp in length
            *   28513   gap of unknown length
            *   28513 29881: contig of 1369 bp in length
            *   29882   gap of unknown length
            *   29882 30600: contig of 719 bp in length
            *   30601   gap of unknown length
            *   30601 31204: contig of 604 bp in length
            *   31205   gap of unknown length
            *   31205 32333: contig of 1129 bp in length
            *   32334   gap of unknown length
            *   32334 32911: contig of 578 bp in length
            *   32912   gap of unknown length
            *   32912 33816: contig of 905 bp in length
            *   33817   gap of unknown length
            *   33817 34778: contig of 962 bp in length
            *   34779   gap of unknown length
            *   34779 36264: contig of 1486 bp in length
            *   36265   gap of unknown length
            *   36265 38613: contig of 2349 bp in length
            *   38614   gap of unknown length
            *   38614 39144: contig of 531 bp in length
            *   39145   gap of unknown length
            *   39145 41224: contig of 2080 bp in length
            *   41225   gap of unknown length
            *   41225 41982: contig of 758 bp in length
            *   41983   gap of unknown length
            *   41983 43378: contig of 1396 bp in length
            *   43379   gap of unknown length
            *   43379 44855: contig of 1478 bp in length
            *   44855   gap of unknown length

```
*   44857   46513: contig of 1657 bp in length
*                  gap of unknown length
*   46514   48278: contig of 1765 bp in length
*                  gap of unknown length
*   48279   49123: contig of 845 bp in length
*                  gap of unknown length
*   49124   50185: contig of 1062 bp in length
*                  gap of unknown length
*   50186   51798: contig of 1613 bp in length
*                  gap of unknown length
*   51799   52794: contig of 996 bp in length
*                  gap of unknown length
*   52795   54154: contig of 1360 bp in length
*                  gap of unknown length
*   54155   56271: contig of 2117 bp in length
*                  gap of unknown length
*   56272   58993: contig of 2722 bp in length
*                  gap of unknown length
*   58994   60268: contig of 1275 bp in length
*                  gap of unknown length
*   60269   62633: contig of 2365 bp in length
*                  gap of unknown length
*   62634   65099: contig of 2466 bp in length
*                  gap of unknown length
*   65100   66776: contig of 1677 bp in length
*                  gap of unknown length
*   66777   68874: contig of 2098 bp in length
*                  gap of unknown length
*   68875   72266: contig of 3392 bp in length
*                  gap of unknown length
*   72267   74691: contig of 2425 bp in length
*                  gap of unknown length
*   74692   77588: contig of 2897 bp in length
*                  gap of unknown length
*   77589   80107: contig of 2519 bp in length
*                  gap of unknown length
*   80108   84825: contig of 4718 bp in length
*                  gap of unknown length
*   84826   87029: contig of 2204 bp in length
*                  gap of unknown length

alignment_scores:
       Quality:       26.00         Length:      26
       Ratio:          1.000        Gaps:         0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AC009127

Align seg 1/1  to: AC009127  .from: 1  to: 186591

   151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProThr 167
       |||||||||||||||||||||||||||||||||||||||||||||||||||
 28398 TGGTTATGGCCGCGGAATCCACCAATTCGTGAACGCAAAGTATTCCCAC 28447

   167 rSerTrpLeuLysIleThrLeuTyrTyrGlu 176
       |||||||||||||||||||||||||||||||
 28448 CAGCCGGCTTAAGAtCACCCTATAtGAA 28475

seq_name: gb_ba2:U32752

seq_documentation_block:
LOCUS       U32752     11833 bp    DNA              BCT       29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 67 of 163 of the complete genome.
ACCESSION   U32752 L42023
VERSION     U32752.1  GI:1573692
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
ORGANISM    Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 11833)
AUTHORS     Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
```

```
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 11833)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 11833)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 11833)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 11833)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The whole genome was shifted by 588 nucleotides for a new start
COMMENT     On Sep 30, 1996 this sequence version replaced gi:1221393.
FEATURES             Location/Qualifiers
     source          1..11833
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            145..939
                     /gene="HI0690"
     CDS             145..939
                     /gene="HI0690"
                     /note="similar to GB:L19201 SP:P11244 GB:M55990 GB:U13915
                     GB:X15054 percent identity: 73.18; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycerol uptake facilitator protein (glpF)"
                     /protein_id="AAC22350.1"
                     /db_xref="GI:1573694"
                     /translation="MDKSLKANCIGEFLGTALLFFGVGCVAALVAGASFGLWEISI
                     MWGMGVALAVYATAGLSGAHLNPAVTIALWKFACFDGKKVIPYIISQMLGAFFRAALV
                     YALYRNVFIDYETVHNIVRGTQESLSLAGFTSYPHPSLSISGGAFAVEFVTAILMAL
                     IMALTDDGNGVPRGPLAPLLIGLIIAVIGGAMGPLTGFAMNPARDFGPKFFAYLAGWG
                     ELALTGGREIPYFIVPMVAPVLGALAGAWLYKKAIGGNLPCNCGCE"
     gene            960..2471
                     /gene="HI0691"
     CDS             960..2471
                     /gene="HI0691"
                     /note="similar to GB:L19201 SP:P08859 GB:M18393 GB:M55990
                     GB:X15054 percent identity: 76.89; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
```

```
                  /product="glycerol kinase (glpK)"
                  /protein_id="AAC22351.1"
                  /db_xref="GI:1573695"
                  /translation="MTDKKYIIALDQGTTSSRAVLLDHNANVVEIAQREFTQIIPRAG
                  WVEHNPMEIWATQSSTLNEVVAKAGITSDEIAAIGITNQRETTIVWEKSTGTPVYNAI
                  VWQCRRTADITDKLKRADGHEEYIRNTTGLVVDPYSFSGTKVKWILDNVEGAREKAERGE
                  LLFGTVDTWLVWKLTQGRVHVTDYTNASRTMLFNIHTKQWDDKMLEILNIPRSMLPEV
                  RNSSEIYGQTNIGGKGGGVRIPVAGIAGDQQAALYGHLCVHAGQAKNTYGTGCFMLLHT
                  GNKAITSKNGLLTTIACNAKGEPEYALEGSVFIAGASIQWLRDELKIVHDSFDSEYFA
                  QKVTDSNGVVVPAFTGLGAPYWDPYARGAIFGLSRGANRNHIVRATLESIAYQTRDV
                  LEAMQSDSGERLQYLRVDGGATNNNFLMQFQADILDVNVERPVVKEVTALGAAYLAGL
                  ATGFWKDLDELRDKARVERTFSPDSDNEKRERRYKGWKKAVKRSLEWAKEDEE"

gene              complement(2546..3013)
                  /gene="HIO692"

CDS               complement(2546..3013)
                  /gene="HIO692"
                  /note="similar to SP:P26972 GB:X52093 GB:X63336 PID:47695
                  percent identity: 74.03; identified by sequence
                  similarity; putative"
                  /codon_start=1
                  /transl_table=11
                  /product="xanthine-guanine phosphoribosyltransferase
                  (gptB)"
                  /protein_id="AAC22352.1"
                  /db_xref="GI:1573696"
                  /translation="MSEKYVTWDMFQMHARRLSERLLPASQWKGIIAVSRGGLFPAA
                  VLARELGLRHIETVCIASYHDHNNQGELQVLHAAQVPNGGEGFIVVDDLVDTGNTARA
                  IRQMYPNAKFVTVFAKPAGAELVDDYVIDIPQNTWIEQPWDLGLITFVPPLSRK"

gene              3297..4121
                  /gene="HIO693"

CDS               3297..4121
                  /gene="HIO693"
                  /note="similar to GB:M68502 SP:P26093 PID:1573696 percent
                  identity: 99.64; identified by sequence similarity;
                  putative"
                  /codon_start=1
                  /transl_table=11
                  /product="lipoprotein E (hel)"
                  /protein_id="AAC22353.1"
                  /db_xref="GI:1573697"
                  /translation="MKTTLKMTALAALSAFVLAGCGSHQMKSEGHANMQLQQQAVLGL
                  NWMQDSGEVKALAYQAYNAAKVAFDHAKVAKGKKKAVVADLDETMLDNSPYAGWQVQN
                  NKPFDGKDWRWDVARQSRAVPGAVEFNNYVNSHGKVFYTVNRKDSTEKSGTIDDMK
                  RLGFNGVEESAFYLKKDKSAKAARFAEIEKQGYEIVLVYGDNLDDFGNTVYGKLNADR
                  RAFVDQNQGKFGKTFIMLPNANYGGWEGGLAEGYFKKDTQGQKARLDAVQAWDGK"

gene              4261..4983
                  /gene="HIO694"

CDS               4261..4983
                  /gene="HIO694"
                  /note="similar to GB:U00096 SP:P75966 PID:1787380 percent
                  identity: 66.29; identified by sequence similarity;
                  putative"
                  /codon_start=1
                  /transl_table=11
                  /product="conserved hypothetical protein"
                  /protein_id="AAC22354.1"
                  /db_xref="GI:1573697"
                  /translation="MMLQAFQNRNHKLMKKNFSAGKLPSKGKSAVNFHRTFKPKLPK
                  TLSLDETKVVLFNKPFDVLTQFTDEQGRATLKDFISIPNVYAAGRLDRDSEGLLILTN
                  NGELQHRLADPKFKTEKTYWVDVEGIPEETDLAQLRKGVELKDGVTRKSAKVRLISEPN
                  LWERNPIRERKNIPTSWLEIKISEGRNRQVRRMTAHIGFPTLRLVRVSMGLLSINGL
                  ENGSFRLLSLLDEIKALFQTVKL"

gene              complement(5109..6080)
                  /gene="HIO695"

CDS               complement(5109..6080)
                  /gene="HIO695"
                  /note="similar to GB:L06129 SP:P29014 PID:147343 GB:U00096
                  PID:1788848 percent identity: 53.62; identified by
                  sequence similarity; putative"
                  /codon_start=1
                  /transl_table=11
                  /product="exopolyphosphatase, putative"
                  /protein_id="AAC22355.1"
                  /db_xref="GI:1573698"
                  /translation="MNDSILEPKHRGNVREIAAIDLGSNSFHMIVARIVNGSIQVLSR
                  LKQKVKLAEGLDENAVLNQEIRTGVNCLALFAERLQGFPMENVNVVGTYTLRRAVNN
                  DEFLRQAAKVPPYPINIISQOTEAKTIYAGVCHTOPEKGRKLVIDIGGGSTEMIIGDD
                  FTPLMAESRHMGCVSFATQFFTDGIISPENFQRARQSAVNKIEDLGLEYRKLGWQSVL
                  GSSGTIKTVAQVIATNLDPNGTITAERLNALIEQTLQAKHFTELNIGLNQDRVDFFV
                  PGLAILSAVFDVFHIQQMRYSDGALREGVIYSLEKNFQVADIRASTA"

gene              complement(6088..9984)
                  /gene="HIO696"

CDS               complement(6088..9984)
                  /gene="HIO696"
                  /note="similar to GB:U14003 SP:P39322 PID:537063 GB:U00096
                  percent identity: 34.87; identified by sequence
                  similarity; putative"
                  /codon_start=1
                  /transl_table=11
                  /product="conserved hypothetical protein"
                  /protein_id="AAC22356.1"
                  /db_xref="GI:1573699"
                  /translation="MTEQIOPSETSPSPEPKPNKHWVRKAVCIGSAVILVPVLGVAG
                  ALSFDAGQKSLIQLVDKMLDSFSVEQVEGGLQNGLVLKNVRYQTAGIETHIAQARIQL
                  DFGCLFSREVCLRDFTLNKPTIAINTALLPPSAPDNSKSGSMKRISLPISINAENLVM
                  QDLSVNIDQTSITLGNFKSAVSLNNEKGLIIAPTEINDISVIAKKLSEVKSEPRAEQP
                  NKPVDWAAIEQSLIPAFLGNVSEIILPFDLHIPEISGKNWQYQAVNEKGETLQSVEMS
                  SLIAQADTVDNQLQLQKLAVESSLGNLSSQGKLQLDGDMPLDLTLKSHLEPLKSDGKE
                  ILPASDVDLTLSGSLKKSTALSLKTKGVLDAELNQVLAQDKMPLNLTLNVAKGQYT
                  FVNTMTPLKINDVTLKLTGDLLNYHAELKGDVAGMNYIPASQVELNADGKLYEVTVNK
                  LGIDSLDGKSEFVGNANWKNGANWDIQADLEKMNIAFFVPVMPATLSGKLHSRGFAGS
                  QGWQVEVPVADLNGMLSAKPISLKGSATLNQNVLLTVPDLQIKYGENYLKASGVLDDH
                  SDFALDINAPNLRGLWSDLKGRVKGRVAISGQITTPNLDLDLTSSNLHLQGFQLAKAS
                  IKGHNNASLSSGKLNIKAEQLHYGNIKLHLDLDLSGDEQNHKLILKSQGEPVAAN
                  LQINGHFDRTLEQWKGTISOVKFETPIGDVKSNQAIAVSYDNKQTQANIASHCWQNTD
                  VELCFPQAFNAGKQGNIPFQFKRVNLDLVNKLIEQNSLKGNLQVQGNVAWFTDKPFQF
                  TANVDGNHLAFSQKIDYRIFKLYIPKLTLNADIQNNNLVLKTDIVHNQGRIVGDIHL
                  NDLAKNRQLGGTLAIERLNLSIANQLLTSGESVNGEVVSKLSFGGNLEKPLLNGDFNI
                  RNIRTKLKSMPVNITDGDIALRFNDNRSTLQGKIKTVDSHLNLTGRANWANIEHWTTE
                  LNAQANNFNVDIPSMAKLRFSPNITIKANPKELNLSGTVDIPWARIKIDSLPDTAEPV
                  SEDEVILNGPHKSKEELIKREFAAKTKSGMEIRSDLRINIGKDVSLDAYGLKTNLDGL
                  LSVKQDKGNLGLFGQINLTKGRYASFGQDLLIRKGLISFSGQATQPTLNIEAIRNPET
                  MEDSKITAGVRVIGIADSPEVTIFSEPSKPQDQALSYLLTGRSLESSGEVGSTGSVGA
                  ALIGLGISKSGKLVGSIGEVFGIODLNLGTSGVDKSKVTVSGNITNRLQIKYGVLF
                  DGLAEVTLRYRLMPQLYFQSVSSTNQVFDLLYKFEF"

gene              complement(9994..11730)
```

```
alignment_scores:
     Quality:          15.00               Length:            15
       Ratio:           1.000                 Gaps:             0
Percent Similarity:    100.000       Percent Identity:    100.000

alignment_block:
US-09-252-691-7056 x U32752

Align seg 1/1   to: U32752   from: 1   to: 11833

74  ValTyrAlaAlaGlyArgLeuAspAspSerGluGlyLeuLeu 88
      |||||||||||||||||||||||||||||||||||||||||
4510 GTTTATGCGGCAGGTCGATTAGATAGAGACAGTGAAGGATTACTG 4554

seq_name: gb_ba2:SLE250721

seq_documentation_block:
LOCUS       SLE250721      5476 bp     DNA              BCT        07-JAN-2000
DEFINITION  Synechococcus leopoliensis crtQ gene (partial), dxr gene, ORF2,
            ORF3, ORF4 and ORF5 (partial), strain SAUG 1402-1.
ACCESSION   AJ250721
VERSION     AJ250721.1   GI:6689328
KEYWORDS    crtQ gene; deoxyxylulose 5'-phosphate reductoisomerase; dxr gene;
            ORF2; ORF3; ORF4; ORF5; zeta carotene desaturase.
SOURCE      Synechococcus leopoliensis
  ORGANISM  Synechococcus leopoliensis
            Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
REFERENCE   1  (bases 1 to 5476)
AUTHORS     Miller,B.H., Heuser,T. and Zimmer,W.L.
```

```
TITLE     Characterization of the functional involvement of adeoxyxylulose
          5-phosphate reductoisomerase gene habouringlocus of the
          Synechococcus leopoliensis genome in isoprenoidbiosynthesis
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 5476)
AUTHORS   Zimmer,W.
TITLE     Direct Submission
JOURNAL   Submitted (29-OCT-1999) Zimmer W., Atmosphaerische Umweltforschung,
          Fraunhofer Institut, Kreuzeckbahnstrasse 19,
          Garmisch-Partenkirchen, D-82467, GERMANY
FEATURES          Location/Qualifiers
     source       1..5476
                  /organism="Synechococcus leopoliensis"
                  /strain="SAUG 1402-1"
                  /db_xref="taxon:32047"
                  /country="Germany"
     gene         complement(1..856)
                  /gene="crtQ"
     CDS          complement(<1..856)
                  /gene="crtQ"
                  /codon_start=1
                  /transl_table=11
                  /product="zeta carotene desaturase"
                  /protein_id="CAB65434.1"
                  /db_xref="GI:6689329"
                  /translation="MRVAIVGAGLAGLAAAIDLVDAGHQVAIYDSRPFVGGKVGSWID
ADGNHIEMGLHVFFNYANLFALMRKVGAFENLLPKAHTHFINKGGEVGELDFRFPI
GAPPNGLKAFFTTSQLTWLDKLQNALALGTSPLVRGILDYEGAMKIIRALDRISFADW
FRSHGGSEGSLKRMWNPIAYALGFIDTENISARCMLTVFQMFAAKTEASKLNLLAGSP
AEYLHKPILDYIQARGATLHLRRRVREIYETYTENGQTVVTGLQIADGDAVERVEADVY
LAACDVPGI"
     gene         988..2207
                  /gene="dxr"
     RBS          988..992
                  /gene="dxr"
     CDS          999..2207
                  /gene="dxr"
                  /function="involved in isoprenoid biosynthesis"
                  /codon_start=1
                  /transl_table=11
                  /product="deoxyxylulose 5-phosphate reductoisomerase"
                  /protein_id="CAB65435.1"
                  /db_xref="GI:6689330"
                  /translation="MKAVTLLGSTGSIGTQTLDILEQYPDRFRLVGLAAGRNVALLSE
QIRRHRPEIVAIQDAAQLSELQAAIADLDNPPLLTGEAGVTEVARYGDAEIVVTGIV
GCAGLLPTIAAIEAGKDIALANKETLIRAGPVVLPLLQKHGVTITPADSEHSAIFQCI
QGLSTHADFRPAQVVAGLRRILLTASGGAFRDWPVERLSQVTVADALKHPNWSMGRKI
TVDSATLMNKGLEVIEAHYLFGLDYDYIDVIHPQSIIHSLIELEDTSVLAQLGWPDM
RLPLYALSWPDRLSTQWSALDLVKAAGSLEFREPDHAKYPCMDLAYAARGKGTMPAV
LNAANEQAVALFLEEQIHFSDIPRLIERACDRHQTEWQQQPSLDDILAYDAWARQFVQ
ASYQSLESVV"
     RBS          2206..2209
                  /note="ORF2"
     CDS          2217..2816
                  /note="ORF2"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAB65436.1"
                  /db_xref="GI:6689331"
                  /translation="MVYRYLLFHKPYDAVCQFSPSDRPDQQTLKDYIDVPEVYPVGRL
DRDSEGLLLLTNNGALQHRLCHPRFGHDRTYWCVEREPTERAALQALRQGVQIODYRT
RPAKVQRLDDPQIPERDPPIRFRKTVPTAWLALTLQEGRNRQVRRMTAAVGHPTLRLI
RAAIATHPDQPPLQLQGLAPGTWRDLTAIECRRLESLLR"
     RBS          2827..2830
                  /note="ORF3"
     CDS          2846..3757
                  /note="ORF3"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAB65437.1"
                  /db_xref="GI:6689332"
                  /translation="MTAGMSESIPSEASFVAAADERTAMLSLAQMEVAATTDVGM
RRSQNEDFYYHYQLEHLETPSGTSVQAKGLFVLCDGMGGHARGEIASLEAARVFKQE
IEAQWTTELPSREALRSAVLAANQALFELNQAAAVSGSDRMGTTIVAILLQGTRVAII
HVGDSRAYRYSRLTGLDQLTRDHEVGQLAIAQGVDVPDVAYARPEAHQLTQGLVEI
WLDPEVRFLEVREDTVFLLCSDGLSDFDLVEEVTEHVAPLQTRSLQYGIQQLVEI
ANEAGGHDNITVVAVRLRPQLGPVI"
     CDS          complement(3791..4384)
                  /note="ORF4"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAB65438.1"
                  /db_xref="GI:6689333"
                  /translation="MLGRDQOPLITADGLTQAEQLGKQWQQTCCWRPQAIWTSPLQRAQ
QTAIALSQGQTPVIPLLVQPALTEIDLGCLTGLTWPEAEQRYPDLCQQLLRDRHWRPI
PEAETLSDCRMRAEQVCTQLIHTEADLWIVISHGGFLQHLLAALLGCDRSWGLDWPLLG
AIDLSLAHEDWDTAGRDRWNTSLWRWQILHWPSRSAI"
     RBS          complement(4398..4402)
                  /note="ORF4"
     CDS          complement(4410..>5476)
                  /note="ORF5"
                  /codon_start=3
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAB65439.1"
                  /db_xref="GI:6689334"
                  /translation="ACRSLLQEPFCLVLSDRLSFLVILVPYPSQPSCYFSQTPEVLAE
AVRLLAFTAHHFPALQPELARALHRFPLANVAETWAPFWAAVLQQWPIVAASDRPS
TEEVSDRPRSTVGADLELLQAIAHEVKTPLATIRTLTRSLLRRVDLPAQARKRLEAI
DLECSEQIDRFGLFFRAVELENRESQALLARTSLTEVEAETVPWRWKQAERRSLALD
LQLPERMPAVVSDPMLDQALTSLVERFTRNLPAGSRIEIVVSLAGAQLKRLRTATEL
VSDPSEQPRTAPRSLGQLLTVQPETGSLSLKLEVSKNLFQALGGKLTVRRRSPQASEV
ITVFLPVEPSDDATADSPR"
     BASE COUNT     1135 a   1592 c   1626 g   1123 t
     ORIGIN

alignment_scores:
          Quality:    12.00          Length:       12
          Ratio:       1.000         Gaps:          0
          Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x SLE250721

Align seg 1/1  to: SLE250721  from: 1    to: 5476

     176 GluGlyArgAsnArgGlnValArgArgMetThrAla 187
         |||||||||||||||||||||||||||||||||||
    2631 GAAGGACGCAATCGCCAGTACGGCGGATGACCGCT 2666

seq_name: gb_bal:D90899

seq_documentation_block:
LOCUS       D90899       133859 bp   DNA                BCT       07-FEB-1999
DEFINITION  Synechocystis sp. PCC6803 complete genome, 1/27, 1-133859.
ACCESSION   D90899 AB001339
VERSION     D90899.1  GI:1651650
KEYWORDS    pNIL34; 2nd component required for LlaI restriction activity; ABC
            transporter; DNA ligase; GDP-D-mannose dehydratase; HtaR suppressor
            protein; N utilization substance protein; NifS protein;
            UDP-N-acetyl-D-mannosaminuronic acid transferase; a negative
            regulator of pho regulon; adenylate kinase; anthranilate synthetase
            alpha-subunit; arginine decarboxylase; biopolymer transport ExbB
            protein; cysteine synthase; dTDP-6-deoxy-L-mannose-dehydrogenase;
            dihydrodipicolinate reductase; dimethyladenosine transferase; drug
            sensory protein A; erythrocyte band 7 integral membrane protein,
            protein 7.2B, stomatin; ferric aerobactin receptor;
            ferrichrome-iron receptor; geranylgeranyl pyrophosphate synthase;
            glycogen (starch) synthase; guanylate kinase; hydrogenase
            isoenzymes formation protein HypD; integrase-recombinase protein;
            iron(III) dicitrate transport system permease protein FecB;
            viron(III) dicitrate transport system permease protein FecC;
```

```
             iron(III) dicitrate transport system permease protein FecD;
             iron(III) dicitrate transport system permease protein FecE;
             iron(III) dicitrate-binding periplasmic protein; leader peptidase
             I; malic enzyme; mutator MutT protein; peptide methionine sulfoxide
             reductase; phosphoglycerate mutase; phosphoribosylformyl
             glycinamidine synthetase II; photosystem I subunit II; photosystem
             II 13 kD protein; photosystem II D1 protein; phycocyanin alpha
             phycocyanobilin lyase CpcF; polyribonucleotide
             nucleotidyltransferase; prenyltransferase; regulatory protein PchR;
             regulatory protein PcrR; ribonuclease E; ribonuclease HII;
             thioredoxin M; transposase; type 4 prepilin peptidase; zeaxanthin
             glucosyl transferase; PNII34.
SOURCE       Synechocystis sp. (strain:PCC6803) DNA.
  ORGANISM   Synechocystis sp.
             Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE    1  (bases 1 to 133859)
  AUTHORS    Tabata,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi
             Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure
             2; 1523-3, Yanauchino, Kisarazu, Chiba 292, Japan
             (E-mail:tabata@kazusa.or.jp, Tel:+81-438-52-3933,
             Fax:+81-438-52-3934)
REFERENCE    2  (sites)
  AUTHORS    Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,
             Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T.,
             Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
             Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
             Yamada,M., Yasuda,M. and Tabata,S.
  TITLE      Sequence analysis of the genome of the unicellular cyanobacterium
             Synechocystis sp. strain PCC6803. II. Sequence determination of the
             entire genome and assignment of potential protein-coding regions
  JOURNAL    DNA Res. 3 (3), 109-136 (1996)
  MEDLINE    97061201
COMMENT      Potential protein coding regions were assigned on the basis of
             similarity search of the ORFs and GeneMark analysis.
FEATURES             Location/Qualifiers
     source          1..133859
                     /organism="Synechocystis sp."
                     /strain="PCC6803"
                     /db_xref="taxon:1143"
     gene            1..772
                     /gene="ispB"
     CDS             <1..772
                     /gene="ispB"
                     /note="ORF_ID:slr0611"
                     /codon_start=2
                     /transl_table=11
                     /product="prenyltransferase"
                     /protein_id="BAA16579.1"
                     /db_xref="GI:1651651"
                     /translation="ARHRRLAEITEMIHTASLVHDDVVDEADLRRNVPTVNSLFDNRV
                     AVLAGDFLFAQSSWYLANLDNLEVVKLLSEVIRDFAEGEILQSINRFDTDTDLETYLE
                     KSYFKTASLIANSAKAAGVLSDAPRDVCDHLYEYGKHLGLAFQIVDDILDFTSPTEVL
                     GKPAGSDLISGNITAPALFAMEKYPLLGKLIEREFAQAGDLEQALELVEQGDGIRRSR
                     ELAANQAQLARQHLSVLEMSAPRESLLELVDYVLGRLH"
     CDS             937..1494
                     /note="ORF_ID:slr0612"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16580.1"
                     /db_xref="GI:1651652"
                     /translation="MGRLDQDSEGLLLITSNGKLQHRLAHREFAHQRTYFAQVEGSPT
                     DEDLEPLRRGIFRADYFPRPPAIAKIITEPDFPPRNPPIRYRASIPTSWLSITILTEGRN
                     RQVRRTAAVGFTLRLVRVQIQVTRSPQQGKGKSAATWCLITLEGLSPGQWRPLTPW
                     EENFCQQLITGNPNGPWQKKFGDRR"
     CDS             1577..2098
                     /note="ORF_ID:slr0613"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16581.1"
                     /db_xref="GI:1651653"
                     /translation="MSYLIAVVANRIAEEAYTTLEQAGFAQKNLTIGTGKTADEF
                     GLVDPKKQAIKRAKLMAIWLVPFGFFAAGYCFNLITGLSTLDWAGDPGNHIVGGLLGAI
                     GGTMGGSFFVGGGVGLSFGSGDSLPYRNLLQAGKYLVVVAGGELQKQRATNLLRPLNPE
                     YLQGYTAPDEAFV"
     CDS             complement(2172..2873)
                     /note="ORF_ID:sll0558"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16582.1"
                     /db_xref="GI:1651654"
                     /translation="MSDNLTELSQQLHDASEKKQLTAIAALAEMGEGGOGILLDYIAK
                     NVPLEKPVLAVGNVYQTLRNLEQETITTQLQRNYPTGIFPLQSAQGIDYLPIQEALGS
                     QDFETADEITRDKLCELAGPGASQRQWLYFTEVEKFPALDLHTINALWLHSNGNFGF
                     SVQRRLWLASGKEFTKLWPKIGWKSGNVWTRWPKGFTWDLSAPQGHLPLLNQLRGVRV
                     AESLYRHPVWSQYGW"
     gene            complement(3192..4268)
                     /gene="pNIL34"
     CDS             complement(3192..4268)
                     /gene="pNIL34"
                     /note="ORF_ID:sll1214"
                     /codon_start=1
                     /transl_table=11
                     /product="PNIL34"
                     /protein_id="BAA16583.1"
                     /db_xref="GI:1651655"
                     /translation="MVNTLEKPGFDEIRPGVKTPAKETILTPRFYTTDFEMAKMDIS
                     PNEDELRAILEEFRVDYNRHHFVRNESFNKSWDHIDGEKRQLFVEFLERSCTAEFSGF
                     LLYKELGRRLKNKNPLLAECFNLMSRDEARHAGFLNKAMSDFNLSLDLGFLTKSRRYT
                     FFKPKFIFYATYLSEKIGYWRYITIYRHLEKNPNDCIYPIFEFFENWCQDENRHGDFF
                     DAIMRAQPHTLNDWKAKLWCRFFLLSVFATMYLNDTQRADFYACLGLEARSYDKEVIE
                     KTNETAGRVFPIILDVNNPEFYNRLETCVSNNEQLRAIDASGAPGVIKALRKLPIFAS
                     NGWQFIKLYLMKPIAVDQLAGAVR"
     CDS             complement(4451..5389)
                     /note="ORF_ID:sll1213"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAA16584.1"
                     /db_xref="GI:1651656"
                     /translation="MLSLENQRILVTGGAGFLGKQVVAQLIAAGGDRPKITIPRSKDC
                     DLRQASACERAVENQDIVIHLAAHVGGIGLNRERPAELFYDNLMGVQLIHAAHQAGV
                     KKFVCVGTICAYPKFTPVPFKEEDLWNGYPEETNAPYGVAKKALLVQLESYRLQYGFN
                     GIYLLPVNLYGPEDNFDPRSSHVIPALIHKVYEAQKAGOKQLPVWGDGSPTREFLYST
                     DAARGIVMGTQAYDKADPVNLGTNFEISIKDLTELICELMEFEGDIIWETDQPNGQPR
                     RCLDTTKAKAEFGEFAQVSLREGLKNTIDWYRQQRS"
     gene            complement(5534..56511)
                     /gene="rfbD"
     CDS             complement(5534..6622)
                     /gene="rfbD"
                     /note="ORF_ID:sll1212"
                     /codon_start=1
                     /transl_table=11
                     /product="GDP-D-mannose dehydratase"
                     /protein_id="BAA16585.1"
                     /db_xref="GI:1651657"
                     /translation="MSKSKVVLLTGITGQDGSYLSELLEKGVQVHGIRRSTFNTD
                     RIDHLYVDPHDLEAKLRLHYGDLTDGTTLRRILEDVKPTEIYNLGAQSHVRVSFDSPE
                     YTVDSVAMGTLRLLEAIRDYQHRTGIQVRFYQAGSSEMFGKVOEIPOKETTPFYPRSP
                     YACAKVYGHWQTVNYRESVDLFACNGILFNHESPRRGEFVTRKITRAIARIVAGTQK
                     KLYLGNIDSKRDWGYAKDYVRAMWAMLQQEPQDDYVVATGETHVKEFLEIAFGYVNL
                     NWQNYVAFDERYLRPAEVDLLIGDPAKTKAQLGWEPSVTFTELVHLMVEADLAVLGLT
                     SPNQSGRIKELMAQDMAFIRSQNGHAVD"
     gene            7229..8311
                     /gene="psdA2"
     CDS             7229..8311
                     /gene="psbA2"
                     /note="ORF_ID:slr1311"
                     /codon_start=1
                     /transl_table=11
                     /product="photosystem II D1 protein"
                     /protein_id="BAA16586.1"
```

```
MEDLINE     20036896
REFERENCE   2  (bases 1 to 9999)
AUTHORS     White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE       Direct Submission
JOURNAL     Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..9999
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            complement(71..655)
                     /gene="DR0891"
     CDS             complement(71..655)
                     /gene="DR0891"
                     /note="similar to GP:1498492 percent identity: 61.93;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-binding response regulator"
                     /protein_id="AAF10469.1"
                     /db_xref="GI:6458612"
                     /translation="MRLVIADDHPLFRMGLKYALIQQGFDVVAEAADGPSALDACRTY
                     QPDAALLDVKMPGMTGIEVCDRLRLTAPRVVSILITTFAEPAIVQAARAGAKGVVSK
                     ETDPESLARQLREIVAHPEVDRLPQVDVPRLTPRESEVLPLLARGFSNKEIARDLGVS
                     PDTVKDHLARLYSKLNAGDRTEAVSRARSIGLLS"
     gene            complement(652..1761)
                     /gene="DR0892"
     CDS             complement(652..1761)
                     /gene="DR0892"
                     /note="similar to PID:2245589 percent identity: 47.14;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sensor histidine kinase"
                     /protein_id="AAF10470.1"
                     /db_xref="GI:6458613"
                     /translation="MTRMTPPAAPRAAAAAPPERRPARLAAVRAGRSRPPLQTLQFT
                     LVVSLLAFLPNLVLALASGAGSLSWSAAAWMLLVGLLSAVIGYVLSGALLRPLHRLEL
                     EVSECDFAQFHADDPREIRSLRRAFGSLMTRLGTEQTRRNAFMATLVHDLKTPLIATG
                     HLTRALTELPLPDDEKRAVGQEIQTETERLLALVQQMADAHRFERDDVQLHLLAPGDLR
                     EVAEQAARRLRPQAEARGLSHVSGSGQAEIDAAALERAVGNLTENALRYARHEVRLT
                     AAPGLLRVQDDGPGLSASLDDLAQPFNAQPAVIAGQOYTAGTAGIGLFIVRRIAEAHG
                     GTLRYGREPPRGRLPAAPGAFPLITFTLVLPEVLP"
     gene            complement(1922..2617)
                     /gene="DR0893"
     CDS             complement(1922..2617)
                     /gene="DR0893"
                     /note="similar to GB:U00096 PID:1787004 percent identity:
                     66.98; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF10471.1"
                     /db_xref="GI:6458614"
                     /translation="MVKSMQQIAMTQQKTLDQVRTFMARTYSWMAAGLALTAGVAYLT
                     AQNEGLAMQVASLRLPLMLALQALVVFVLSMFAQRLSAVAGALFVGAAALTGFSAL
                     LFAYSPAAVITAFAVSAGTFGLMSVAGFVIKKDLSAMGRFFLFAVLGLVVAMLVNLFV
                     GSSALSLGISMIGVPLFAGLTAYDTQMLRNLALSGISGEQAERASINGALALYLDFIN
                     IFLPLLNIGNSRD"
     gene            complement(2702..4207)
                     /gene="DR0894"
     CDS             complement(2702..4207)
                     /gene="DR0894"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
```

```
     gene            8492..10471
                     /gene="speA"
     CDS             8492..10471
                     /gene="speA"
                     /note="ORF_ID:slr1312"
                     /codon_start=1
                     /transl_table=11
                     /product="arginine decarboxylase"
                     /protein_id="BAA16587.1"
                     /db_xref="GI:1651659"
                     /translation="MGEEPVPADKALGKFKKNASWSIEESEALYRVEAWGAPYFAI
                     NAAGNITVSPNGDRGGSLDLLELVEALRQRKLGLPLLIRFSDILADRLERLNSCFAKA
                     IARYNYPNTYQAVYPVKCNQQRHLVEALVRFGQTSQCGLEAGSKPELMIALATLPPPL
                     DRQDKHTKPLIICNGYKDQDYLETALLAKRLGHRPIIIIEQLRELEWVLHISOQLNIK
                     PMLGVRARLSCQSLKSSEISSGNGDRAKLGLITMPDIVTVIHRLEENNCLDCLKMLHFH
                     LGTQVSDTALIKEAMREASQLYVELVKLGAKMRYLNVGGGLAVDYDGSKTNYPASKNY
                     NMONANDIVAAIQDACELGQVSPPILVSESGRAIMAHQSVLVFDVLGSNQTGFSEPH
                     PPDENAHPLLKNLWECYETITAEQYQEQYHDALQLKTEASSLFNFGYILSLTERGQAEQ
                     IHWACCRKIFETTRQLEYIPEDFQALDRIMTDIYYVNLSVFQSAPESWSLDQLFPILP
                     IHHLNEKPSORVILADLTCDSDGKIDRFIDLWDVKSYLEVHPLENDGNPYYLGMFLVG
                     AYQEIMGNLHNLFGDINVVHIATTPQGYQIESVVRGDTMTEVLGYVQYDSDDLLEGLR
                     RHTELALSNGQITLEESRRLLEDYEQSLRRYTYLS"
     gene            complement(10622..12631)
                     /gene="lig"
     CDS             complement(10622..12631)
                     /gene="lig"
                     /note="ORF_ID:sll1209"
```

```
     gene            8492..10471
                     /gene="speA"
     CDS             8492..10471
                     /gene="speA"
                     /note="ORF_ID:slr1312"
                     /codon_start=1
                     /transl_table=11
```

```
                     /db_xref="GI:1651658"
                     /translation="MTTTLQQRESASLWEQFCQWVTSTNNRIYVGWFGTLMIPTLLFA
                     TTCFIIAFIAAPPVDIGIREPVAGSLLYGNNIISGAVVPSSNAIGLHFYPIWEAASL
                     DEWLYNGGPYQLVFHFLIGIFCYMGRQWELSYRLGMRPWICVAYSAPVSAATAVFLI
                     YPIGQGSFSDGMPLGISGTFNEMIVFQAEHNILMHPFHMLGVAGVFGGSLFSAMHGSL
                     VTSSLVRETTEVESQNYGYKFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLG
                     AWPVIGIWFTAMGVSTMAFNLNGFNFNQSILDSQGRVIGTWADVLNRANIGFEVMHER
                     NAHNFPLDLASEQAPVALTAPAVNG"
```

```
alignment_scores:
     Quality:    12.00        Length:       12
     Ratio:      1.000        Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x D90899
Align seg 1/1  to: D90899  from: 1 to: 133859

   176 GluGlyArgAsnArgGlnValArgArgMetThrAla 187
       ||||||||||||||||||||||||||||||||||||
  1231 GAGGGGCGCAATCGTCAGGTACGTCGAATGACAGCG 1266

seq_name: gb_ba2:AE001942

seq_documentation_block:
LOCUS       AE001942   9999 bp   DNA            BCT   22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 79 of 229 of the complete
            chromosome 1.
ACCESSION   AE001942 AE000513
VERSION     AE001942.1 GI:6458611
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
ORGANISM    Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcus.
REFERENCE   1  (bases 1 to 9999)
AUTHORS     White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE       Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
JOURNAL     Science 286 (5444), 1571-1577 (1999)
```

```
               DVSKSFGNYIRAQLILLVISGISGVGLLLLGVPNALALGLLTGLNLVPLVGMVLAA
               IPALLQALPLGTTKVLIVAGLYFVLNQVAWNVIAPHVMGRTVNLSAASILIAILIGGG
               LAGLPGALLAIPTASLLVLMLRRYWFPSPAYQGDGELAVPMDPVRLPADVEPPQGKSQ
               PE"
     gene      7999. .8538
               /gene="DR0899"
     CDS       7999. .8538
               /gene="DR0899"
               /note="similar to SP:P29253 PID:581811 percent identity:
               67.92; identified by sequence similarity; putative"
               /codon_start=1
               /transl_table=11
               /product="ribonuclease H"
               /protein_id="AAF10474.1"
               /db_xref="GI:6458617"
               /translation="MTRPGRPSARKKPDTSRDLLPIRAGIQPEVPVGGQVVELYSDGA
               CDTTKGHGGWATILRYGEREVLSGNEENTNNRMELRGLLEGLRTLRRPCQVKVITD
               SQYLRKAFTDGWILNWQRNGWKTASKEPVKNQDLWEELIELAKVHALTFLWVKGHAGH
               GENERVDELAVLERKKLRK"
     gene      complement(8559. .9008)
               /gene="DR0900"
     CDS       complement(8559. .9008)
               /gene="DR0900"
               /note="identified by Glimmer2; putative"
               /codon_start=1
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="AAF10480.1"
               /db_xref="GI:6458623"
               /translation="MLRAMNIADPLGVPAAPLETWQAMQARLNDADLTGARLILITDR
               QGERPWAGYAALVLGRGAAEEAILSAPAFGPRYGAAGAALAELVRWADSAGIRPTE
               TILGSGDFQRVLAEPDAGEVAALIAATNPADPAITTPQPATEPEWV"
     gene      9071. .9457
               /gene="DR0901"
```

```
alignment_scores:
          Quality:      10.00          Length:        10
          Ratio:         1.000         Gaps:           0
  Percent Similarity:  100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AE001942

Align seg 1/1  to: AE001942  from:  1  to: 9999

    176 GluGlyArgAsnArgGlnValArgArgMet 185
        |||||||||||||||||||||||||||||||
   5227 GAGGGCCGCAACCGTCAGGTGCGCCGCATG 5256

seq_name: gb_bal:BACDIA

seq_documentation_block:
LOCUS       BACDIA       28206 bp    DNA             BCT    26-MAY-1995
DEFINITION  Bacillus subtilis spoVA to serA region.
ACCESSION   L09228
VERSION     L09228.1  GI:410114
KEYWORDS    3-dehydroquinate dehydratase; aroC gene; diaminopimelate
            decarboxylase; lysA gene; penicillin-binding protein;
            peptidyl-prolyl isomerase; phosphoglycerate dehydrogenase; ppiB
            gene; response regulator; response regulator kinase; ribA gene;
            ribB gene; ribD gene; ribG gene; ribH gene; ribT gene; riboflavin
            biosynthesis operon; serA gene; signal peptidase; sipS gene; spoA
            gene; spoVAF gene.
SOURCE      Bacillus subtilis (strain 168, sub_species Marburg) DNA.
ORGANISM    Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 28206)
AUTHORS     Yamamoto,J., Shimizu,M. and Yamane,K.
TITLE       Molecular cloning and analysis of nucleotide sequence of the
            Bacillus subtilis lysA gene region using B. subtilis phage vectors
            and a multi-copy plasmid, PUB110
JOURNAL     Agric. Biol. Chem. 55 (6), 1615-1626 (1991)
```

```
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="AAF10477.1"
               /db_xref="GI:6458620"
               /translation="MIKGKELLGRPVVAVSNGAQVDNVHDVVFDHQGNRVLALLVEEG
               GWLRAAKAVPFGRVQSIGEKAVMIATPEDVVNTGRDPVLKEALAGKTNLIGMTLLTTD
               GQTLGKITDVYFDERSGQVEGYEASGGLFADMTNGRTFIPAPQSIQIGKDSAIVPAGV
               AAAMREQAPGGIRGAFHTAGKSLGGAVQGAADSVRGAYEDTSRSVRENYEQVSANVAE
               AAREGQQQARAAASDLRSRVQGGAEDLREASAERQRDFVVGKVAGQDVRTDNGVLIVA
               KGTTIDEGQARFAEQNGALPALLAAAAGGSVREGVRDLLADPSSDADVVGRRVQEDV
               RSSVGSLIAVQGQIVTPEVAEBRARELGLEERIVAATQAPGLQTPPVPDRAELREQASD
               LRDNLADGLSNVSQNVSQGANELLGRAKSWIGERRDLEDSLDRQEEAYQARIRDAL
               GRPVTRAVLRTDRVLQPGELVTHRAALDEARAAGVLGLLLDSVDGGAPLS"
     gene      4311. .4697
               /gene="DR0895"
     CDS       4311. .4697
               /gene="DR0895"
               /note="identified by Glimmer2; putative"
               /codon_start=1
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="AAF10478.1"
               /db_xref="GI:6458621"
               /translation="MLRAAFWMTALLIVPLGLLLFYFLPGSAAQLIGISPIWLARASGG
               LLLAWGLFQLFAGAQPDGAKVGGLVTGNLLTVATLLPALLKLQTSLPPSLRLVLWVVV
               GWLGLAALALFAAPLGGRGGEAGVR"
     gene      4687. .5460
               /gene="DR0896"
     CDS       4687. .5460
               /gene="DR0896"
               /note="similar to GB:L09228 SP:P35159 PID:410137
               GB:AL009126 percent identity: 61.70; identified by
               sequence similarity; putative"
               /codon_start=1
               /transl_table=11
               /product="ribosomal large subunit pseudouridine synthase
               B"
               /protein_id="AAF10472.1"
               /db_xref="GI:6458615"
               /translation="MSAERLHKRLARAGIASRRAAEELIRAGRVTVNGQTAGLGQGVN
               DTDDVRVDGRLVELTRPETVTYALYKPVGFVTTAHDEYGRRNVLDAMPDVPGLHPVGR
               LDKDSEGLLLLNDGDLTLTLTHPRYGHEKAYRAWTEGREPPTQAELDVIVRGIAMDD
               GPAQALSAAPAEDGAYVVLGEGRNRQVRRMLEALGHPVGRLVRIRVGGLWLGDLNPGE
               YRELGPRDLEQLLHPHKTPRAVWDGAWEQTQDRWGTQYG"
     gene      complement(5992. .6579)
               /gene="DR0897"
     CDS       complement(5992. .6579)
               /gene="DR0897"
               /note="identified by Glimmer2; putative"
               /codon_start=1
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="AAF10479.1"
               /db_xref="GI:6458622"
               /translation="MRVPRRSWPPRRFPPVLLSGAVTLKRMPQPVTALLAAMTPAVL
               ISGAGTLIMSTSTRVGRATDRVRGLTSRFKQLVSAEGQNEPLAREEKRLIIRQLPRLS
               RRTRFLMRALQAFYTSVALLVLTSILLGLDGLFDVNTGSLPLLVALGGSVLAYGALL
               LSYEARYSAATTKEEMRFLEELGQHYAELYADELS"
     gene      6600. .7784
               /gene="DR0898"
     CDS       complement(6600. .7784)
               /gene="DR0898"
               /note="similar to GB:AL009126 percent identity: 56.70;
               identified by sequence similarity; putative"
               /codon_start=1
               /transl_table=11
               /product="conserved hypothetical protein"
               /protein_id="AAF10473.1"
               /db_xref="GI:6458616"
               /translation="MGNALRTVPQLLAFLWSRSYVRLIVYAIRALLLIGLARQAAGL
               ALIRIAYALSYTNPLLLWLEKHRITRPGVAATLIVVLGVIGVLFWSVASQVVSFLQ
               GLPALLNSLPQLLEHLIGKGPGHHNAGISQTANQLTVQYEQVDNLTRNFGPLLARVL
               SPDSALYGQVSGVMNWLGQAGLVLTLAVFFMLDHSKFGRTLLRVFPKEWQPRLLLSE"
```

```
MEDLINE   91345841
REFERENCE 2  (bases 1 to 28206)
AUTHORS   Buchanan,C.E. and Ling,M.L.
TITLE     Isolation and sequence analysis of dacB, which encodes a
          sporulation-specific penicillin-binding protein in Bacillus
          subtilis
JOURNAL   J. Bacteriol. 174 (6), 1717-1725 (1992)
MEDLINE   92193254
REFERENCE 3  (bases 1 to 28206)
AUTHORS   Sorokin,A., Zumstein,E., Azevedo,V., Ehrlich,S.D. and Serror,P.
TITLE     The organization of the Bacillus subtilis 168 chromosome region
          between the spoVA and serA genetic loci, based on sequence data
JOURNAL   Mol. Microbiol. 10 (2), 385-395 (1993)
MEDLINE   95020538
FEATURES             Location/Qualifiers
     source          1..28206
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /sub_species="Marburg"
                     /db_xref="taxon:1423"
     gene            1..1239
                     /gene="spoVAF"
     CDS             <1..1239
                     /gene="spoVAF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA67472.1"
                     /db_xref="GI:410115"
                     /translation="VDIVENRLLNAQVEKVKTLDETTDQVLSGLVAVIVEGAGFAFII
                     DVRSYPGRNPEEPDTEKVVRGARDGFVENIVVNTALLRRIRDERLRVKMTKVGERSK
                     TDLSICYIEDIADPDLVEIVEKEIASIDVDGLTMADKTVEEFIVNQSYNPFFLVRYTE
                     RPDVAANHVLEGHVIIIVDTSPSVIITPTLFHHVQHAEEYRQAPSVGTFLRKVRFFG
                     ILASTLFLPIWFLFVLQPDLLPDNMKFIGLNKDTHIPIILQIFLADIGIEFLRMAAIH
                     TPTALSPAMGLIAAVLIGQIAIEVGLFSPEVILYVSLAAIGFTTPSYELSLATNEPS
                     CPHDTRCFISYKRARHRLYSANYAMASIKSLQTPYLMPLIFPNGKALMQVLVRTAKPG
                     AKVRPSIVHPKNRLRQPTNS"
     conflict        1158..1164
                     /gene="spoVAF"
                     /citation=[1]
     -35_signal      1245..1250
                     /gene="lysA"
     gene            1245..2664
                     /gene="lysA"
                     /note="putative"
     -10_signal      1266..1272
                     /gene="lysA"
     CDS             1345..2664
                     /gene="lysA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="diaminopimelate decarboxylase"
                     /protein_id="AAA67473.1"
                     /db_xref="GI:410117"
                     /translation="MFLHGTSRQNOHGHLEIGGVDALYLAEKYGTPLYVYDVALIRER
                     AKSFKQAFISAGLKAQVAYASKRAFSSVAMIQLAEEEGLSLDVVSGGELYTAVAAGFPA
                     ERIHFHGNNKSREELRMALEHRIGCIVVDNFYEIALLEDLCKETGHSIDVLLRITPGV
                     EAHTHDYTTGQEDSKFGFDLHNGQTERAIEQVLQSEHIQLLGVHCHIGSQIFDTAGF
                     VLAAEKIFKKLDEWRDSYSFVSKVLNLGGFGIRYTEDDEPLHATEVVEKIIEAVKEN
                     ASRYGFDIPEIWIEPGRSLVGDAGTTLYTVGSQKEVPGVRQYVAVDGGMNDNIRPALY
                     QAKYEAAAANRIGEAHDKTVSIAGKCCSGDMLIWDIDLPEVKEGDLLAVFCTGAYGY
                     SMANNYNRIPRPAVVFVENGEAHLVKRETYEDIVKLDLPFKTGVKQ"
     conflict        1610..1616
                     /gene="lysA"
                     /citation=[1]
     conflict        1751..1757
                     /gene="lysA"
                     /citation=[1]
     conflict        2496..2502
                     /gene="lysA"
                     /citation=[1]
     conflict        2504..2510
                     /gene="lysA"

     terminator      2666..2694
     terminator      complement(2666..2694)
                     /note="ORFX19"
     CDS             complement(2693..3565)
                     /note="ORFX19"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA67474.1"
                     /db_xref="GI:410118"
                     /translation="MKKIWIGMLAAAVLLMVPKVSLADAAVGDVIVTLGADLSESDK
                     QKVLDEMNVPDNATTVTVTNKEEHEYLGKYISNAQIGSRAISSSSITIAKKGSGLNVE
                     THNISGITDEMYLNALMTAGVKDAKVYVTAPFEVSGTRALTGLIKAYEVSSDEAISED
                     VKQVANQELVTTSELGDKIGNENAAALIAKIKEEFAKNGVPDNKADIEKQVDDAASDL
                     NVTLTDSQKNQLVSLFNRKMKNADIDWGQVSDQLDKAKDKITKFIESDEGKNFIQKVID
                     FFVSIWNAIVSIFK"
     conflict        3347..3353
                     /citation=[1]
     conflict        3413..3419
                     /citation=[1]
     conflict        3428..3434
                     /citation=[1]
     conflict        3503..3509
                     /citation=[1]
     gene            3819..4290
                     /gene="ppiB"
     CDS             3819..4250
                     /gene="ppiB"
                     /codon_start=1
                     /trans_table=11
                     /product="peptidyl-prolyl isomerase"
                     /protein_id="AAA67475.1"
                     /db_xref="GI:410119"
                     /translation="MSYWIGYRLDGDKIEFELYPEAAPGTVANFEKLANEGFYDGLTF
                     HRVIPGFVSQGGCPHGTGTGGPGYTIKCETEGNPHTHEAGALSMAHAGKDTGGSQFFI
                     VHEPQPHLNGVHTVVFGKVTSGLEFAKNMSNGDVMKEVRVEG"
     conflict        3982..3988
                     /gene="ppiB"
                     /citation=[1]
     conflict        4261..4290
                     /gene="ppiB"
     terminator      4287..4320
                     /note="ORFX2"
     repeat_region
                     /note="inverted repeat"
     CDS             5264..5467
                     /note="ORFX1"
                     /codon_start=1
                     /trans_table=11
                     /protein_id="AAA67476.1"
                     /db_xref="GI:410120"
                     /translation="MQVLTGLLEQYGYAVLFIVLMLELLALLGPEFAGVLVFQEKLN
                     WVLSMLAAGLGSCCGSEHVVLDRV"
     CDS             5445..5834
                     /note="ORFX2"
                     /codon_start=1
                     /trans_table=11
                     /protein_id="AAA67477.1"
                     /db_xref="GI:410121"
                     /translation="MSYWIGYRLGCPAAIOKAFASRGVLLAVIVLTFFWTVFRGQDRKI
                     ELSFLFFLVIGGELYEEGIRRLFHSLHPVHAPLESHVLYPFPSEQTLTAFVIFGFCAY
                     LFVRHSHKTRLQTVAFFCLLSFYSSSG"
     -35_signal      6022..6027
                     /gene="sipS"
                     /note="putative"
     gene            6022..7323
                     /gene="sipS"
     -10_signal      6045..6051
                     /gene="sipS"
                     /note="putative"
     -35_signal      6683..6688
                     /gene="sipS"
                     /note="putative"
     -10_signal      6705..6710
                     /gene="sipS"
```

```
CDS                /note="putative"
                   6740..7294
                   /gene="sipS"
                   /codon_start=1
                   /transl_table=11
                   /product="signal peptidase"
                   /protein_id="AAA67478.1"
                   /db_xref="GI:410122"
                   /translation="MKSENVSKKKSILEWAKAIVIAVVLALLIRNFIFAPYVVDGDSM
                   YPTLHNRERVFNVMTVKYIGEFDRGDIVVLNGDDVHYVKRIIGLPGDTVEMKNDQLYI
                   NGKKVDEPYLAANKKRAKQDGFDHLTDDFGPVKVPDNKYFVMGDNRRNSMDSRNGLGL
                   FTKQIAGTSKFVFYPFNEMRKTN"
terminator         7299..7323
-35_signal         /gene="sipS"
                   7388..7393
                   /note="putative; ORFX4"
-10_signal         7410..7416
                   /note="putative; ORFX4"
CDS                7529..7873
                   /note="ORFX4"
                   /codon_start=1
                   /transl_table=11
                   /protein_id="AAA67479.1"
                   /db_xref="GI:410123"
                   /translation="MGRIKTKITILVLLLLAGGYMYINDIELKDVPTAIGQTLSSE
                   EEBYTIQEYKVTKIDGSEYHGVAENGTKIIFNGKKLNQDLSDIKEGDKIKAYFSKSKR
                   IDGLIKVAKVND"
terminator         7872..7911
                   /note="ORFX4"
-35_signal         7937..7942
-10_signal         7960..7965
protein_bind       8126..8137
                   /note="putative"

alignment_scores:
    Quality:              10.00          Length:        10
    Ratio:              1.000            Gaps:           0
Percent Similarity:  100.000    Percent Identity:  100.000

alignment_block:
US-09-252-691-7056 x BACDIA

Align seg 1/1  to: BACDIA  from: 1  to: 28206

  176 GluGlyArgAsnArgGlnValArgArgMet 185
      ||||||||||||||||||||||||||||||
17989 GAGGGACGCAACAGACAAGTGCGCCGCCATG 18018

seq_name: gb_bal:BSUB0013

seq_documentation_block:
LOCUS       BSUB0013     218470 bp    DNA              BCT     26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 13 of 21): from 2395261
            to 2613730.
ACCESSION   299116 AL009126
VERSION     299116.1  GI:2634723
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1  (bases 1 to 218470)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,Y., Fujita,Y., Fuma,S., Gallizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hulio,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 218470)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1..218470
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     terminator      complement(7..18)
                     /gene="cmk"
     gene            complement(7..696)
                     /gene="cmk"
     CDS             complement(22..696)
                     /gene="cmk"
                     /EC_number="2.7.4.14"
                     /note="alternate gene name: JofC, YpfC"
                     /codon_start=1
                     /transl_table=11
                     /product="cytidylate kinase"
                     /protein_id="CAB14221.1"
                     /db_xref="GI:2634724"
                     /db_xref="SWISS-PROT:P38493"
                     /translation="MEKKLSIAIDGPAAAGKSTVAKIVAEKKSYIYIDTGAMYRAITY
AALQENVDLTDEEKLAELLKRTDIELITTKDGQKFVNGDVTEAIRTDEISNQVSIA
AKHRSVREEMVKRQQLGEKGGVVMDGRDIGTHVLPNAEVKIFLLASVERAKRRYEE
NVKKGFDVNFETLIEEIARRDKLDSEREVSPLRKAEDALEIDTTSLSIQEVADKILEA
VEQKSR"
     gene            complement(775..951)
                     /gene="ypfB"
     CDS             complement(775..951)
                     /gene="ypfB"
                     /note="unknown"
                     /note="alternate gene name: JofB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB14222.1"
                     /db_xref="GI:2634725"
                     /db_xref="SWISS-PROT:P38492"
                     /translation="MKTFERLLIKLLFIQAILLGVQFLFHYQHIEPYVSKVIQYEGV
DKMEENNRIETFKH"
     gene            complement(996..1649)
                     /gene="ypfA"
     CDS             complement(996..1649)
                     /gene="ypfA"
                     /function="unknown"
```

```
            /note="alternate gene name: jofA"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14223.1"
            /db_xref="GI:2634726"
            /db_xref="SWISS-PROT:P38491"
            /translation="MIEIGENVLLEYIEENELKKAKSKAVSIENNELLIAYPVDVVTG
            RTVILHNDMEVTVEFVGKDEVPYRFISRIKGKVKDKLQMICLEMPPREKMKRIQRRQY
            VRTDAVLDVQIQPGNEEEIRTLSYNISAGGIAVVLADGLSFQSGESLRLIIRLPEEEH
            TRQIETEAVVRIRMKVALDDGQQALLQYCIRRQLNKRKRARME"
gene        complement(1711..3094)
            /gene="ypeB"
terminator  complement(1711..1733)
            /gene="ypeB"
CDS         complement(1742..3094)
            /gene="ypeB"
            /function="unknown"
            /note="alternate gene name: joeB, yzuA; similar to
            hypothetical proteins"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14224.1"
            /db_xref="GI:2634727"
            /db_xref="SWISS-PROT:P38490"
            /translation="MIRGILIAVLGIAIVGTGYWGYKEHQERDAVLLHAENNYQRAFH
            ELTYQVDQLHDKIGTTLAMNSQKSLSPALIDVWRITSEAHNSVSQLPLTLMPFNKTEE
            LLSKIGDFSYKTSVRDLDQKPLDKNEYTSLNKLYQQSEDIQNELRHVQHLVMSKNLRW
            MDVEMALASDEKQSDNTIINSFKTVEKNVGAFSTGTDLPGPSFTSTTKKEEKGFSHLKGK
            QISEQEAKQIAERFAPDNYSIKVVKSGKKTNRDVYISISMKDPDHKAVIYMDITKKGG
            HPVYLIQNREVKDQKISLNDGSNRALAFLKKNGFETDDLEIDESAQYDKIGVFSYYPV
            ENKVRMYPEAIRMKVALDDGEVYGWGTWAVRNFQDQFGLKEVDGLVGAKTKQTLICKSK
            VRETRLALITNELGQEVLCYEMLGTIENDTFRMYINAKDGSEEKVEKLKNAEPIYKDL"
gene        complement(3129..4046)
            /gene="sleB"
CDS         complement(3129..4046)
            /gene="sleB"
            /note="alternate gene name: ypeA"
            /codon_start=1
            /transl_table=11
            /product="spore cortex-lytic enzyme"
            /protein_id="CAB14225.1"
            /db_xref="GI:2634728"
            /db_xref="SWISS-PROT:P50739"
            /translation="MKSKGSIMACLILFSFTITTFINTETISAFSNQVIQRGATGDDV
            VELQARLQYNGYYNGIKDGVYGWGTWAVRNFQDQFGLKEVDGLVGAKTKQTLICKSK
            YYREYVMEQLNKGNTFTHYGKIPLKIYQTKPSKAATQKARQOAEARQKQPAEKTQKPK
            ANANKQQNNTPAKARKQDAVAANMPGFSNNDIRLLAQAVVGEARGEPYEGQVAIAAV
            ILNRLNSPLFPNSVAGVIFEPLAFTAVADGQIYMQPNETAREAVLDAINGWDPSEEAL
            YYFNPDTATSPWIWGRPQIKRIGKHIFCE"
gene        complement(4162..4841)
            /gene="ypdC"
terminator  complement(4162..4182)
            /gene="ypdC"
CDS         complement(4185..4841)
            /gene="ypdC"
            /function="unknown"
            /note="similar to hypothetical proteins"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14226.1"
            /db_xref="GI:2634729"
            /db_xref="SWISS-PROT:P50738"
            /translation="MFAIISAGIAPGIALLSYFYLKDQYDNEPVHMVLRSFLGVVLV
            FPIMFIQYVLEKENVGGGSFFVSFLSSGFLEESLKWFILMISVYPHAHFDEHYDGIVY
            GASVSLGFATLENILYLIGHGVEHAFVRALLPVSCHALIGVIMGFYLGKARFSADKAR
            VKWLTLSLVVPSLLHGSYDFIITALSNWIYMLPFMVFLWWFGLRKAKKARSVNMMQV
            "
gene        complement(4929..5935)
            /gene="ypdA"
terminator  complement(4929..4953)
            /gene="ypdA"
CDS         complement(4961..5935)
            /gene="ypdA"
```

```
            /gene="ypdA"
            /function="unknown"
            /note="similar to thioredoxin reductase"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14227.1"
            /db_xref="GI:2634730"
            /db_xref="SWISS-PROT:P50736"
            /translation="MIQGKAIIIGGGPCGLSAAIHLKGIGIDALVIEKGNVVNSIYNY
            PHQTFFSSEKELEIGDVAFITENRKPVRIQALSYYREVKRKNIRVNAFEMVRKVTK
            TQNNTFVIETSKETYTPYCIIATGYDHPNYMGVPGEDLPKVFHYFKEGHPYFDKDV
            VVIGGKNSSVDAELVKSGARVTVLVRGNEYSPSIKPWILPEFEALVRNGTIRMEFG
            ACVEKITENEVVFRSGEKELITIKNDFVFAMTGYHPDHQFLEKIGVEIDKETGRPFFN
            EETMETNVEGVFIAGVIAAGNNANEIFIENGRFHGGHIAAEIAKRENH"
gene        complement(5995..7324)
            /gene="ypcA"
terminator  complement(5995..6016)
            /gene="ypcA"
CDS         complement(6044..7324)
            /gene="ypcA"
            /function="unknown"
            /note="similar to:glutamate dehydrogenase"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14228.1"
            /db_xref="GI:2634731"
            /db_xref="SWISS-PROT:P50735"
            /translation="MAADRNTGHTEEDKLDVLKSTOTVIHKALEKLGYPEEVELLKE
            PMRLLTVKIPVRMDDGSVKIFTGYRAHNDSVGPTKGGIRFHPNTEKEVKAVKALSIW
            MSLKCGIDLPYGGGKGGIVCDPRDMSFRELERLSRGYVRAISOIVGPTKDVAPDVF
            TNSQIMAWMDEYSRIDEFNSPGFITGKPLVLGGSHGRESATAKGVTICIKEAAKKRG
            IDIKGARVVVQGFGNAGSYLAKFMHDAGAKVVGISDAYGGLYDPEGLDIDYLLDRRDS"

alignment_scores:
            Quality:       10.00          Length:       10
            Ratio:          1.000         Gaps:          0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x BSUB0013/rev   ..
Align seg 1/1 to reverse of: BSUB0013  from: 1 to: 218470

   176 GluGlyArgAsnArgGlnValArgArgMet 185
       ||||||||||||||||||||||||||||||
 25594 GAGGACGGCCAACAGACAAGTGCGCCGCATG 25565

seq_name: gb_pr3:HSN113A11

seq_documentation_block:
LOCUS       HSN113A11     22633 bp    DNA              PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone LL22NC03-113A11 on chromosome 22
            Contains the 3' part of the gene for KIAA0045, ESTs and GSSs,
            complete sequence.
ACCESSION   Z83856
VERSION     Z83856.1  GI:2995261
KEYWORDS    HTG; KIAA0045.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 22633)
AUTHORS     McLaren,S.
TITLE       Direct Submission
JOURNAL     Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 28, 1998 this sequence version replaced gi:2558587.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
```

only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
LL22NC03-113A11 is from the human chromosome 22-specific cosmid library LL22NC03, constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550 under the auspices of the National Laboratory Gene Library Project sponsored by the US Department of Energy. The source of the flow sorted chromosomes was a human/hamster hybrid containing chromosomes Y, 22 and 9. VECTOR: lawrist16
IMPORTANT: This sequence is not the entire insert of clone LL22NC03-113A11 It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone LL22NC03-113A11 is at 1 in this sequence. The true right end of clone RP1-180M12 is at 100 in this sequence. The start of this sequence overlaps with sequence 282190
The end of this sequence overlaps with sequence 282248.

```
FEATURES             Location/Qualifiers
source          1..22633
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22"
                /clone="LL22NC03-113A11"
                /clone_lib="LL22NC03"
repeat_region   16..106
                /note="MIR repeat: matches 56. .147 of consensus"
repeat_region   1593..1910
                /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   2252..2565
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   2676..2974
                /note="AluSq repeat: matches 1. .299 of consensus"
repeat_region   3015..3179
                /note="AluJ repeat: matches 3. .127 of consensus"
repeat_region   3180..3480
                /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region   3481..3590
                /note="AluJ repeat: matches 127..243 of consensus"
mRNA            join(<3961. .4021,8202. .8284,8423. .9230)
                /gene="cN113A11.1"
                /note="match: cDNAs: Em:AB014545; match: ESTs: Em:AA793897
                Em:AA650024 Em:AA259014 Em:AI040181 Em:AA215487
                Em:AI358529 Em:H55161 Em:AA976016 Em:AA215486 Em:AA832576
                Em:AA927357."
                /evidence=not_experimental
                /product="cN113A11.1 (KIAA0645)"
gene            3961. .9230
                /gene="cN113A11.1"
CDS             join(<3961. .4021,8202. .8284,8423. .8715)
                /gene="cN113A11.1"
                /note="match: proteins: Tr:075140"
                /codon_start=3
                /evidence=not_experimental
                /product="cN113A11.1 (KIAA0645)"
                /protein_id="CAB62971.1"
                /db_xref="GI:6572403"
                /translation="FEPETYWDRMHLFQEAIAHRFGFVQDKYSASAFNFPAENKPQYI
                HVTGTVFLQLPYSKRKFSGQQRRRNSTSTNQMFCEERVGYNWAYNTMLTKTWRSS
                ATGDEKFADRLLKDFTDFCINRDNRLVTFWTSCLEKMHASAP"
repeat_region   4628. .4671
                /note="11 copies 4 mer tata 88 conserved"
repeat_region   5355. .5946
                /note="296 copies 2 mer cc 56 conserved"
repeat_region   5397. .5492
                /note="4 copies 24 mer 94 conserved"
repeat_region   5506. .5879
                /note="17 copies 22 mer 58 conserved"
repeat_region   5576. .5851
                /note="12 copies 23 mer 68 conserved"
unsure          5608. .5612
                /gene="cN113A11.1"
                /note="Tandem repeat"
misc_feature    5608. .5612
                /gene="cN113A11.1"
                /note="Tandem repeat; Forced join; Restriction data
                estimate of gap 200bp"
misc_feature    5613. .5618
                /gene="cN113A11.1"
                /note="Tandem repeat; Single clone region"
unsure          5613. .5618
                /gene="cN113A11.1"
                /note="Tandem repeat"
repeat_region   5740. .5811
                /note="3 copies 24 mer 94 conserved"
repeat_region   5889. .5980
                /note="4 copies 23 mer 81 conserved"
misc_feature    6042. .6445
                /gene="cN113A11.1"
                /note="match: GSS: Em:AQ494872"
repeat_region   6268. .6477
                /note="5 copies 42 mer 90 conserved"
repeat_region   6680. .6953
                /note="AluSg1 repeat: matches 1. .270 of consensus"
repeat_region   7399. .7517
                /note="AluSg/x repeat: matches 183. .301 of consensus"
repeat_region   7539. .7713
                /note="AluSg/x repeat: matches 122. .296 of consensus"
repeat_region   8842. .8877
                /note="18 copies 2 mer tt 80 conserved"
repeat_region   8845. .8876
                /note="8 copies 4 mer tttg 96 conserved"
misc_feature    complement(9032. .9230)
                /note="match: STS: Em:G38618"
polyA_signal    9211. .9216
                /gene="cN113A11.1"
polyA_site      9230
misc_feature    9351. .9611
                /gene="cN113A11.1"
                /note="match: GSS: Em:AQ263906"
repeat_region   9590. .9906
                /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   10041. .10529
                /note="L2 repeat: matches 1922. .2482 of consensus"
repeat_region   10630. .10939
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   10948. .11026
                /note="L2 repeat: matches 2622. .2710 of consensus"
repeat_region   11070. .11363
                /note="AlurB8 repeat: matches 1. .308 of consensus"
repeat_region   11669. .11834
                /note="AluJo repeat: matches 1. .124 of consensus"
repeat_region   11835. .12129
                /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region   12130. .12284
                /note="AluJo repeat: matches 124. .292 of consensus"
repeat_region   12891. .13019
                /note="AluSg repeat: matches 1. .135 of consensus"
repeat_region   13020. .13156
                /note="AluSx repeat: matches 1. .138 of consensus"
repeat_region   13157. .13460
                /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region   13461. .13623
```

```
repeat_region    /note="AluSx repeat: matches 138. .298 of consensus"
                 13837. .14135
repeat_region    /note="AluSx repeat: matches 1. .299 of consensus"
                 15037. .15236
repeat_region    /note="AluJo repeat: matches 84. .302 of consensus"
                 1523?. .15526
repeat_region    /note="AluSq repeat: matches 1. .292 of consensus"
                 15527. .15599
repeat_region    /note="AluJo repeat: matches 65. .84 of consensus"
                 15600. .15827
repeat_region    /note="AluY repeat: matches 75. .300 of consensus"
                 15828. .15887
repeat_region    /note="AluJo repeat: matches 1. .65 of consensus"
                 15905. .16202
repeat_region    /note="AluSx repeat: matches 1. .305 of consensus"
                 16203. .16303
repeat_region    /note="MER20 repeat: matches 101. .218 of consensus"
                 16581. .16715
misc_feature     /note="2 repeat: matches 2600. .2742 of consensus"
                 16718. .17120
                 /note="match: GSS: Em:AQ532885"
misc_feature     /note="match: GSS: Em:AQ835686"
                 16722. .17065
repeat_region    /note="AluY repeat: matches 1. .311 of consensus"
                 16870. .17176
repeat_region    /note="AluY repeat: matches 1. .296 of consensus"
                 17194. .17504
repeat_region    /note="AluSx repeat: matches 1. .296 of consensus"
                 18173. .18467
repeat_region    /note="AluSx repeat: matches 1. .296 of consensus"
                 18755. .19049
repeat_region    /note="AluJb repeat: matches 1. .303 of consensus"
```

```
alignment_scores:
    Quality:             9.00          Length:            9
    Ratio:               1.000         Gaps:              0
Percent Similarity:    100.000         Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x HSN113All

Align seg 1/1  to: HSN113All  from: 1  to: 22633

   98 AlaArgLeuThrGlnProGlyLysArg 106
      |||||||||||||||||||||||||||
21259 GCCAGTTAACGCAGCCTGGGAAAAGG 21285

seq_name: gb_bal:SHGCPIR
```

```
seq_documentation_block:
LOCUS       SHGCPIR   107379 bp    DNA            BCT       16-AUG-1996
DEFINITION  S.hygroscopicus gene cluster for polyketide immunosuppressant
            rapamycin.
ACCESSION   X86780
VERSION     X86780.1  GI:987088
KEYWORDS    ABC-transporter; antibiotic transport complexes regulator;
            cholesterol oxidase regulator; cystathione synthase; Cytochrome
            P450; helix-turn-helix protein; ketoreductase/dehydrogenase; lysine
            cyclodeaminase; membrane transport protein; methyltransferase;
            monosaccharide transporter; orfD gene; orfDD gene; orfE gene; orfF
            gene; orfG gene; orfH gene; orfR gene; orfU gene; orfV gene; orfG
            gene; orfW gene; orfX gene; orfY gene; orfz gene; orfzz gene;
            polyketide immunosuppressant rapamycin; polyketide synthase;
            pteridine-dependent dioxygenase; rapA gene; rapB gene; rapC gene;
            rapI gene; rapJ gene; rapK gene; rapL gene; rapM gene; rapN gene;
            rapO gene; rapP gene; rapQ gene; rapT gene; response regulator;
            sensory protein kinase.
SOURCE      Streptomyces hygroscopicus.
  ORGANISM  Streptomyces hygroscopicus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 107379)
  AUTHORS   Schwecke,T., Aparicio,J.F., Molnar,I., Khaw,L.E.,
            Haydock,S.F., Oliynyk,M., Caffrey,P., Cortes,J., Lester,J.B.,
            Boehm,G.A., Staunton,J. and Leadlay,P.F.
  TITLE     The biosynthetic gene cluster for the polyketide immunosuppressant
            rapamycin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7839-7843 (1995)
  MEDLINE   95372374
REFERENCE   2 (bases 1 to 107379)
  AUTHORS   Schwecke,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-1995) T. Schwecke, University of Cambridge, Dept
            of Biochemistry, Tennis Court Road, Cambridge CB2 1QW, UK
REFERENCE   3 (bases 1 to 107379)
  AUTHORS   Molnar,I., Aparicio,J.F., Haydock,S.F., Khaw,L.E., Schwecke,T.,
            Konig,A., Staunton,J. and Leadlay,P.F.
  TITLE     Organisation of the biosynthetic gene cluster for rapamycin in
            Streptomyces hygroscopicus: analysis of genes flanking the
            polyketide synthase
  JOURNAL   Gene 169 (1), 1-7 (1996)
  MEDLINE   96186895
REFERENCE   4 (bases 1 to 107379)
  AUTHORS   Aparicio,J.F., Molnar,I., Schwecke,T., Konig,A., Haydock,S.F.,
            Khaw,L.E., Staunton,J. and Leadlay,P.F.
  TITLE     Organization of the biosynthetic gene cluster for rapamycin in
            Streptomyces hygroscopicus: analysis of the enzymatic domains in
            the modular polyketide synthase
  JOURNAL   Gene 169 (1), 9-16 (1996)
  MEDLINE   96186896
REFERENCE   5 (bases 1 to 107379)
  AUTHORS   Haydock,S.F., Aparicio,J.F., Molnar,I., Schwecke,T., Koenig,A.,
            Marsden,A.F.A., Galloway,I.S., Staunton,J., Leadlay,P.F.
            Boehm,G.A., Staunton,J. and Leadlay,P.F.
  TITLE     Divergent sequence motifs correlated with the substrate specificity
            of (methyl)malonyl-CoA: acyl carrier protein transacylase domains
            in modular polyketide synthases
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..107379
                     /organism="Streptomyces hygroscopicus"
                     /strain="NRRL 5491"
                     /db_xref="taxon:1912"
     misc_feature    1..107379
                     /note="byosynthetic gene cluster"
                     /product="polyketide immunosuppressant rapamycin"
     CDS             complement(<1. .871)
                     /gene="orfZz"
                     /codon_start=1
                     /transl_table=11
                     /product="monosaccharide transporter"
                     /protein_id="CAA60449.1"
                     /db_xref="SPTREMBL:Q54286"
                     /db_xref="GI:987089"
                     /translation="MAPRPSVRSVADVVALIDGLGKITGRAQIIWFLVFGGLFLDAYS
                     NAALSAGLGPMTTQMELTSTQVSILTATAPALALFNPLGGWLATRIGRVPPLLIAKL
                     FAIAGALLAAFAGDFTVVWLGRVLVGVAYGMDFAVAMALLAEYTPAKLGGRLNLWQAV
                     WYIATTSNLVLALVFFNLDVGADIWRWSGSAAVVAVALLGQWLMKESPTWLAGKG
                     RLGEAITNLDKIYGIKAVAGTPDEMTRTPAEAPVIGLRQAGRLFRGEYLRPTILSSVI
                     SLGQSMQYFAVGWY"
     gene            complement(1. .871)
                     /gene="orfZz"
     gene            complement(1150. .2319)
                     /gene="orfZz"
     CDS             complement(1150. .2319)
                     /gene="orfZ"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA60450.1"
                     /db_xref="GI:987090"
                     /db_xref="SPTREMBL:Q54287"
                     /translation="MKFLYDDESFSFEALRAAGYAAYGGADLGEVLVTCRQIPEGDEE
                     AWSAQWAATAARIERIGRDALAAGHRVSAREALLRASNYRTADFYRRENPAADAESA
                     RLAKASQQTFADAAPLLDTPARALRIPYEDTTLPGYLFLVDDSGTPRPTVLFHGGYDS
                     TLEEDYLALAAGALRRGYNVIAFDGPGQGSTVREQLHFPPDWEAVVTPVVDFALTLP
                     EVDAEQLVLMGMSLGGYLAARAAAFEHRVAACVLYDGVVDFHAVIAATAGRAASTPGG
                     MEALMAQNTMVRWVVRNGRWTFGVSGVDELVRAAEPYTMAGIADRITCPTLVLEAEND"
```

```
                  QFFKGQPQRLFDELTCKELISFREDEGAGEHCHEGAMFLFHQRTFDWLDTVLAG"
gene              2422..3036
                  /gene="orfY"
CDS               2422..3036
                  /gene="orfY"
                  /codon_start=1
                  /transl_table=11
                  /product="regulator of antibiotic transport complexes"
                  /protein_id="CAA60451.1"
                  /db_xref="GI:987091"
                  /db_xref="SPTREMBL:Q54288"
                  /translation="MTAQQKKGRPATGGAALRQRVTEAITEAAFAELADAGYARMSME
                  SVARRAGVGKAALYRRWPSKQAMVTELIRGKVTDTLPPTATGALRTDLRELLTTFRG
                  QLANPLLARIAAGLLAEASHDDALAEGLYTGVTAPRRAAAHAILRGAIDRGELPPGLD
                  LDLGTDLLIAPLAFRVLVIQGRSDDEYLELTLTNAIEAALRAAVR"
gene              complement(3056..3763)
                  /gene="orfX"
CDS               complement(3056..3763)
                  /gene="orfX"
                  /codon_start=1
                  /transl_table=11
                  /product="ABC-transporter"
                  /protein_id="CAA60452.1"
                  /db_xref="GI:987092"
                  /db_xref="SPTREMBL:Q54289"
                  /translation="MTSATAAPPVLRLAGVSHTYSGQRRRTAVLKDIDYAFERGTFYT
                  ILGPSGSGKTILLSLASGLDTPTKGTITFEQDLITELGLGRYRNQHAATIFQQYNLLT
                  YLTALQNVTCAMEITGTKPTTGVSRKARALHLLEKIGLDKAMATRNVLQLSGGQQQRVA
                  IARALACDVDILFADEPTGNLDEDTAAGIIGTFRELAHEQDKCVVVSHSQQLAAQSD
                  RVLTLRKGKLTEQVNAR"
gene              complement(3770..5149)
                  /gene="orfW"
CDS               complement(3770..5149)
                  /gene="orfW"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAA60453.1"
                  /db_xref="GI:987093"
                  /db_xref="SPTREMBL:Q54290"
                  /translation="MNFVKRAGLSLWSRKGRTLITLATFLVISVMVLAGVLINDATAQ
                  AEQDARRSVGAEVNLEADLGQLGGGSELRAPRIDAATVDKVGALPQVRKYTYSIWDRV
                  LLKGGPQLADGGPKEPMGPGGTVGMGVLDTSLLPDFRSGKFTLLSGEHLTAADKGKKR
                  LLIEERLAKENKLGVGDTITLTGNDEKTTADFTVAGVYRDPSPATEPDSEYGISPANM
                  LYGTVGGLSALSSERNGPLQAGKATFLLQDADTGQAFKDEAKKVAGPALDGFKLDAND
                  KAIRQMTGPLKSISTTATLAMWLMGLAGAAVLTLLVNLAVKQRHKEYGVLLAMGEKKR
                  KLIAQQALEILVVGALAIGLSSLFAPRLTQSAGQALLGREATSAQHKIDSWQGPPSGS
                  TGLGEGIDPNDKPVEDADPIDKITVRLDPADLATLGAIGLGIGLLATAVPAGAVLRLS
                  PRTILTKGK"
gene              complement(5184..6497)
                  /gene="orfV"
CDS               complement(5184..6497)
                  /gene="orfV"
                  /codon_start=1
                  /transl_table=11
                  /product="membrane transport protein"
                  /protein_id="CAA60454.1"
                  /db_xref="GI:987094"
                  /db_xref="SPTREMBL:Q54291"
                  /translation="MNFVKRAGMSLGARKSKSAALLAIFVICTLLLGGFLLQAAAAR
                  QEAGAQRTIGVDVTVRKKGLTAVLADRLGASDLVYRYNPELPVRAGARGFTPLTSDAP
                  KPGDAGAGAGPAGAKKEQGPLAVNGVRDGSMLLPFSYGSTKITAGRGITPEDAG
                  RNAAVIEQRLADRNHLKVGDTIHVRSVDDKETVPLQVVGVFQNPTQDPARWTAPHKLP
                  GNALYVPMGTAQRLGPGNVAVSEAVYKIGSPEQAERLHTEAERLLGGDGFDFRVNDRA
                  YKDQVRPIQRVGAFAGLIVWVIALAGTLIIGLIVALQIRERRAEEGVLLAMGEKRWKL
                  IGQHAVEVAAVALPAIALAALAGSLAGQPAGDTILGRQDGKPAAAAAAGVPDAEIAP
                  PTVRVEPATVGKVAGIGLGISLVATVIPGIGILRLHPRSLLTDTE"
gene              6703..7365
                  /gene="orfR"
CDS               6703..7365
                  /gene="orfR"
                  /codon_start=1
                  /transl_table=11
                  /product="response regulator"
```

```
                  /protein_id="CAA60455.1"
                  /db_xref="GI:987095"
                  /db_xref="SPTREMBL:Q54292"
                  /translation="MRVLVAEDHRVLARTIATGLRRQAMAVDVAAADEAERMCLLTA
                  YDVLILDRDLPVLSGDEVCRRLRELQDPPRILMLTAAGELTDKIYGLTTLGADDYLAK
                  PFDFSELVARVRTLSRRAPKQPPTVLRFASITLDETRREVSVDGRPLELTPKELAVLH
                  LLLVAGGAPVSHDELVRTVWDEHLDPRTSAVRATVSRLGRGKLGDPGLILADKGEGYRL
                  CD"
gene              7356..8555
                  /gene="orfS"
CDS               7356..8555
                  /gene="orfS"
                  /codon_start=1
                  /transl_table=11
                  /product="sensory protein kinase"
                  /protein_id="CAA60456.1"
                  /db_xref="GI:987096"
                  /db_xref="SPTREMBL:Q54293"
                  /translation="MRLNPWALLSRLPFRARLTAAFSALFLIAGIALLAFVVLARHG
                  TEQQAQGISVTYGDVPSGSGAPMGPVRPTRPNLTQRAPGDVAMFQKIDQTVRAVQDTA
                  LRQMVLWSAVGLLAMALLAGVLGWHLAGRALRPVASMTETARRISEQSLHQRLALTGP
                  DDELHRLADFFDTMLDRLEKSFESQRRFVANASHELKTPLTVQRTSLQVGLADPLPEG
                  LADVREDLLTANHEAEQLINGLLLLARSDRGLEKTQTMDVAAIVRLVSTGLTPLATKN
                  GVRIDLDADPLAVPGDPVLLRHLITNLVRNAIQYNHPGGHVRIRLDTPTVTVNTGP
                  HVSPEQIPDLFEPFHRLDGDRTATTGHGLGLSIAHSIANAHHATTTAQPGEGGLTLT
                  LRFPCPR"
gene              complement(8730..9332)
                  /gene="orfU"
CDS               complement(8730..9332)
                  /gene="orfU"

alignment_scores:
        Quality:       9.00        Length:        9
        Ratio:         1.000       Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x SHGCPIR

Align seg 1/1 to: SHGCPIR   from: 1   to: 107379

   94 GlyValLeuGlnAlaArgLeuThrGln 102
      |||||||||||||||||||||||||||
64357 GGTGTCCTCCAGGCCCGCCTCACGCAG 64383

seq_name: gb_htg4:AC011098

seq_documentation_block:
LOCUS       AC011098   158329 bp    DNA              HTG    19-OCT-1999
DEFINITION  Homo sapiens clone 2_L_1, *** SEQUENCING IN PROGRESS ***, 14
            unordered pieces.
ACCESSION   AC011098
VERSION     AC011098.1  GI:6006241
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158329)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone 2_L_1
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 158329)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Macdonald,P., Marquis,N.,
            Lehoczky,J., Lieu,C., Locke,K., McKernan,K., McLaughlin,J., Meldrim,J.,
            McEwan,P., McGurk,A., McKernan,K.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
```

```
               Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Tirrell,A., Vassilev,H., Vo,A., Wheeler,J., Wu,X.,
               Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        All repeats were identified using RepeatMasker: Smit, A.F.A. &
               Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html.
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 14 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *         1      140: contig of 140 bp in length
               *                     gap of unknown length
               *       141     4575: contig of 4435 bp in length
               *                     gap of unknown length
               *      4576    11055: contig of 6480 bp in length
               *                     gap of unknown length
               *     11056    19208: contig of 8153 bp in length
               *                     gap of unknown length
               *     19209    27069: contig of 7861 bp in length
               *                     gap of unknown length
               *     27070    35519: contig of 8450 bp in length
               *                     gap of unknown length
               *     35520    41379: contig of 5860 bp in length
               *                     gap of unknown length
               *     41380    49140: contig of 7761 bp in length
               *                     gap of unknown length
               *     49141    58881: contig of 9741 bp in length
               *                     gap of unknown length
               *     58882    69901: contig of 11020 bp in length
               *                     gap of unknown length
               *     69902    81612: contig of 11711 bp in length
               *                     gap of unknown length
               *     81613    95329: contig of 13717 bp in length
               *                     gap of unknown length
               *     95330   124361: contig of 29032 bp in length
               *                     gap of unknown length
               *    124362   158329: contig of 33968 bp in length.
FEATURES             Location/Qualifiers
     source          1..158329
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2_L_1"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    47572 a  32248 c  30585 g  47923 t    1 others
ORIGIN
```

```
alignment_scores:
        Quality:        9.00            Length:         9
        Ratio:          1.000           Gaps:           0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AC011098/rev

Align seq 1/1 to reverse of: AC011098  from: 1  to: 158329

     98 AlaArgLeuThrGlnProGlyLysArg 106
        |||||||||||||||||||||||||||
  97207 GCTAGACTAACACACAGCCTGGGAAAGA 97181
```

THIS PAGE BLANK (USPTO)

```
OM of: US-09-252-691-7056 to: N_Geneseq_36:*    out_format : pfs

Date: Jun 10, 2000  4:02 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL-frame+-p2n.model -DEV-xlp
  -Q=/cgn2.1/USPTO_spool/US09252691/runat_05062000_101738_1291/app_query.fasta.1
  -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=oligo.rng -GAPOP=4.500
  -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEX=0.000
  -QGAPOP=4.500 -XGAPOP=60.000 -XGAPEXT=60.000
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
  -TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
  -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
  -MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
  -THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 60.370000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:V74878	-	10.00	163.81	0.1595	676	! Staphylococcus aureus contig SE
N_Geneseq_36:X20509	+	8.00	105.72	274.42	12838	! Polynucleotide sequence from
N_Geneseq_36:V69432	+	7.00	128.49	14.79	46	! Human heat shock protein (hsp70B
N_Geneseq_36:T63894	+	7.00	128.49	14.79	46	! Human hsp70B gene TFIID binding
N_Geneseq_36:X17182	+	7.00	128.49	14.79	46	! Test sequence from human hsp70B
N_Geneseq_36:V21143	+	7.00	115.16	81.74	286	! Human hsp-70B promoter. Spatial
N_Geneseq_36:V49571	+	7.00	111.91	124.13	447	! Human lymphoma cell line U937
N_Geneseq_36:V24918	-	7.00	110.02	158.12	579	! H. pylori cytoplasmic protein C
N_Geneseq_36:T01289	-	7.00	109.60	166.70	613	! Rhipicephalus appendiculatus pr
N_Geneseq_36:V24773	+	7.00	109.47	169.59	624	! H. pylori ORF 04gel12l0orf1. He
N_Geneseq_36:X21163	-	7.00	109.28	173.91	641	! Polynucleotide sequence from th
N_Geneseq_36:X15049	-	7.00	109.13	177.21	654	! DNA encoding a GST-MIG fusion p
N_Geneseq_36:X15057	+	7.00	109.13	177.21	654	! DNA encoding GST-MIG fusion pro
N_Geneseq_36:V83108	+	7.00	109.12	177.46	655	! Human C-type lectin-1 (CTL-1) c
N_Geneseq_36:V86546	+	7.00	108.99	180.50	667	! Wilson disease exon 5/intron bo
N_Geneseq_36:V49570	+	7.00	108.67	188.09	697	! Human lymphoma cell line U937
N_Geneseq_36:V54641	-	7.00	108.67	188.09	697	! Nucleotide sequence encoding a
N_Geneseq_36:V54649	-	7.00	107.79	210.47	786	! Serpin proteinase inhibitor-2
N_Geneseq_36:T73345	-	7.00	105.61	278.18	1059	! DNA encoding lipase modulator
N_Geneseq_36:V74709	+	7.00	105.07	298.28	1141	! Staphylococcus aureus contig S
N_Geneseq_36:V52017	+	7.00	104.82	308.05	1181	! DNA encoding a human MCG18 prot
N_Geneseq_36:V25104	+	7.00	104.72	311.95	1197	! H. pylori cytoplasmic protein
N_Geneseq_36:V34569	+	7.00	104.64	315.37	1211	! Human IL-6 antisense oligonucle
N_Geneseq_36:005610	+	7.00	104.23	332.15	1280	! nifH gene and promoter region
N_Geneseq_36:N50059	+	7.00	104.23	332.33	1281	! Sequence of the Fe-protein sub
N_Geneseq_36:Q50979	+	7.00	104.00	342.33	1322	! CRABP-II gene 5'-end. Human ce
N_Geneseq_36:N82138	+	7.00	103.47	366.26	1421	! Gene encoding tropoelastin. cD
N_Geneseq_36:V15846	+	7.00	103.24	377.34	1467	! Nucleic acid encoding CrmA. Co
N_Geneseq_36:T33568	+	7.00	103.23	377.58	1468	! Cowpox virus CrmA gene. New nu
N_Geneseq_36:T72795	-	7.00	102.46	416.67	1631	! Beta-1 integrin modulator cDNA
N_Geneseq_36:T62311	-	7.00	102.39	420.74	1648	! Rat amelin-1 gene. Enamel matr
N_Geneseq_36:V29052	+	7.00	101.99	442.67	1740	! Alcohol and/or aldehyde dehydr
N_Geneseq_36:Q22824	+	7.00	101.82	452.42	1781	! Sequence encoding ovine interl
N_Geneseq_36:X33292	-	7.00	101.64	462.87	1825	! M. grisea PTH2 gene coding seq
N_Geneseq_36:T62310	-	7.00	101.20	489.87	1939	! Rat amelin-1 gene. 'Enamel m
N_Geneseq_36:T73350	+	7.00	100.87	510.88	2028	! Lipase gene expression cons
N_Geneseq_36:V61572	-	7.00	100.72	521.24	2072	! Human Kv potassium channel
N_Geneseq_36:V61571	-	7.00	100.61	528.77	2104	! Human Kv potassium channel
N_Geneseq_36:V61573	-	7.00	100.61	528.77	2104	! Human Kv potassium channel
N_Geneseq_36:Q67062	-	7.00	100.50	536.05	2135	! Rabphilin-3A cDNA. Target p
N_Geneseq_36:Q48170	+	7.00	100.26	552.71	2206	! Sequence encoding a 90K tum
N_Geneseq_36:Q46888	-	7.00	100.26	552.71	2206	! Sequence encoding an immuno
N_Geneseq_36:Q46161	-	7.00	100.26	552.71	2206	! DNA encoding tumour associa
N_Geneseq_36:Q02147	-	7.00	100.21	556.46	2222	! RuBPCase gen
N_Geneseq_36:T85144	-	7.00	100.20	557.16	2225	! Human Immune Cell Cytokine-
N_Geneseq_36:T02367	+	7.00	99.08	643.18	2594	! Melon aconitase cDNA clone
N_Geneseq_36:T66421	-	7.00	98.21	718.97	2922	! Human P100 protein coding s
N_Geneseq_36:X03030	+	7.00	98.18	721.73	2934	! Human IL-1ra BAC contiguous
N_Geneseq_36:X32291	-	7.00	98.10	729.09	2966	! M. grisea PTH2 gene sequenc
N_Geneseq_36:052536	+	7.00	98.04	735.30	2993	! Encodes Subtilisin-like ser
N_Geneseq_36:V56991	-	7.00	97.28	810.21	3320	! Maize starch phosphorylase
N_Geneseq_36:V36957	+	7.00	96.98	842.13	3460	! Nucleotide sequence of huma
N_Geneseq_36:V57515	+	7.00	96.86	854.87	3516	! Zcytor7 cytokine receptor e
N_Geneseq_36:X32295	+	7.00	96.23	927.23	3835	! M. grisea PTH1 gene sequen
N_Geneseq_36:X19491	-	7.00	95.33	1.0e+03	4337	! Human secreted protein clon
N_Geneseq_36:V04633	+	7.00	94.76	1.1e+03	4692	! Porcine phosphoinositide 30
N_Geneseq_36:V74099	+	7.00	94.76	1.1e+03	4692	! Porcine G-protein regulated
N_Geneseq_36:Q29703	+	7.00	94.11	1.2e+03	5125	! IRS-1. Purified nucleic aci
N_Geneseq_36:T68952	-	7.00	94.08	1.2e+03	5150	! Blackcurrant fruit-specific
N_Geneseq_36:V08418	-	7.00	93.06	1.4e+03	5924	! HCMV strain AD169 genomic f
N_Geneseq_36:Q75977	-	7.00	92.78	1.4e+03	6151	! Insulin receptor substrate-
N_Geneseq_36:X55977	+	7.00	91.80	1.6e+03	7036	! Human factor-related antise
N_Geneseq_36:X33180	-	7.00	91.05	1.8e+03	7797	! Cowpox virus bsr full lengt
N_Geneseq_36:X55272	+	7.00	91.05	1.8e+03	7803	! Human receptor-related anti
N_Geneseq_36:V33476	+	7.00	88.09	2.6e+03	11703	! Sindbis virus clone TR339
N_Geneseq_36:X20501	-	7.00	86.66	3.2e+03	14244	! Polynucleotide sequence fr
N_Geneseq_36:Q40480	+	7.00	85.86	3.5e+03	15894	! Attenuated measles virus s
N_Geneseq_36:V18264	+	7.00	85.86	3.5e+03	15894	! Measles virus Edmonston is
N_Geneseq_36:V18268	-	7.00	85.86	3.5e+03	15894	! Measles virus Rubeovax vac
N_Geneseq_36:V18269	-	7.00	85.86	3.5e+03	15894	! Measles virus Moraten vacc
N_Geneseq_36:V18270	-	7.00	85.86	3.5e+03	15894	! Measles virus Zagreb vacci
N_Geneseq_36:V18271	-	7.00	85.86	3.5e+03	15894	! Measles virus AIK-C vaccin
N_Geneseq_36:V18265	+	7.00	85.86	3.5e+03	15894	! Measles virus 1977 isolate
N_Geneseq_36:Q48171	-	7.00	85.51	3.7e+03	16656	! Eukaryotic layered vector
N_Geneseq_36:T30787	-	7.00	85.51	3.7e+03	16656	! Alphavirus-based eukaryoti
N_Geneseq_36:V42364	-	7.00	85.51	3.7e+03	16656	! Representative eukaryotic
N_Geneseq_36:Q86154	-	7.00	85.51	3.7e+03	16656	! Representative eukaryotic
N_Geneseq_36:V70684	-	7.00	85.51	3.7e+03	16656	! Representative eukaryotic
N_Geneseq_36:X20525	-	7.00	82.64	5.3e+03	24699	! Polynucleotide sequence fr
N_Geneseq_36:T08126	-	7.00	82.07	5.7e+03	26698	! Mouse syndecan-1 gene. New
N_Geneseq_36:Q67902	-	7.00	82.07	5.7e+03	26700	! Syndecan gene. Syndecan st
N_Geneseq_36:V15946	+	7.00	82.07	5.7e+03	26700	! Mouse syndecan gene sequen
N_Geneseq_36:T76260	-	7.00	82.07	5.7e+03	26700	! Mouse syndecan-1 encoding
N_Geneseq_36:X23520	+	7.00	78.18	9.3e+03	45546	! Human kidney aminopeptidas
N_Geneseq_36:T58840_4	+	7.00	71.74	2.1e+04	110000	+ ! Continuation (5 of 6) of
N_Geneseq_36:V30458_3	+	7.00	71.74	2.1e+04	110000	+ ! Continuation (4 of 6) of
N_Geneseq_36:V30459_3	+	7.00	71.74	2.1e+04	110000	+ ! Continuation (4 of 6) of
N_Geneseq_36:X53491	-	6.00	71.42	2.2e+04	114955	! Human adenosine A1 recept
N_Geneseq_36:X54050	+	6.00	116.64	67.66	19	! Human IL-6 receptor antisense
N_Geneseq_36:X19598	+	6.00	116.64	67.66	19	! Human IL-6 receptor antisense
N_Geneseq_36:V91548	+	6.00	116.26	70.99	20	! PCR primer for human globin s
N_Geneseq_36:Q48866	-	6.00	115.57	77.61	20	! Human globin gene primer RS-4
N_Geneseq_36:T47900	-	6.00	114.63	87.47	22	! Nucleotide sequence of the CT
N_Geneseq_36:V31644	-	6.00	114.07	94.00	25	! Cylindrocladium floridanum sp
N_Geneseq_36:V34649	-	6.00	113.55	100.50	27	! Human genomic DNA target sequ
N_Geneseq_36:X19599	+	6.00	113.30	103.74	27	! Human genomic DNA target sequ
N_Geneseq_36:X19606	+	6.00	113.30	103.74	29	! Human C-raf hammerhead riboz
N_Geneseq_36:X38786	+	6.00	113.07	106.97	30	! Human C-raf hammerhead riboz
N_Geneseq_36:Q66689	-	6.00	112.83	110.20	30	! PCR primer IVS-1 ntl10-M to d
N_Geneseq_36:X19332	-	6.00	112.61	113.42	30	! PCR primer IVS-1 ntl10-M to d
N_Geneseq_36:Q97741	+	6.00	111.97	123.04	36	! Neutral protease npr gene pri
N_Geneseq_36:T52591	+	6.00	111.78	126.23	37	! Human genomic DNA target sequ
N_Geneseq_36:Q42536	+	6.00				! Human genomic DNA polymorphic
					31	! Primer C2 to amplify SalI-Stu
					32	! Human Fc-epsilon-R1-alpha PCR
					33	! Primer P283 to subclone T.gon
				111.97	36	! Human ICAM hammerhead ribozym
				126.23	37	! Epidermin structural gene epi

Left table:

ID	Strand		Score	Value	Len	Description
N_Geneseq_36:T76275	-	6.00	111.39	132.60	39	Human neutrophil elastase antise
N_Geneseq_36:X54071	-	6.00	111.39	132.60	39	Neutrophil elastase (Medullasin)
N_Geneseq_36:Q69433	+	6.00	110.19	154.75	46	Human heat shock protein (hsp70B
N_Geneseq_36:X17183	+	6.00	110.19	154.75	46	Human hsp70B gene TFIID binding
N_Geneseq_36:Q50128	+	6.00	109.88	161.03	48	Test sequence from human hsp70B
N_Geneseq_36:Q89014	+	6.00	109.88	161.03	48	Ruthenium-labelled, elongated ba
N_Geneseq_36:T74685	+	6.00	109.43	170.43	51	VEGF 2'-NH2-RNA nucleic acid lig
N_Geneseq_36:T74687	+	6.00	109.43	170.43	51	SAC69 used in construction of GC
N_Geneseq_36:Q23186	+	6.00	109.29	173.56	52	SAC75 used in construction of GC
N_Geneseq_36:T29881	+	6.00	109.15	176.68	53	HPV probe 39. Compns. for ampli
N_Geneseq_36:V76532	+	6.00	109.02	179.79	54	Human papillomavirus detection S
N_Geneseq_36:Q22524	+	6.00	108.75	186.02	56	Staphylococcus aureus contig SEQ
N_Geneseq_36:X17695	+	6.00	108.62	189.12	57	RNA-RNA recombination construct
N_Geneseq_36:Q22524	+	6.00	107.23	226.14	69	pAD-CMV1 primer EBI-2585. O-glyc
N_Geneseq_36:Q20741	+	6.00	107.23	226.14	69	pAD-CMV1 primer EBI-2585. O-glyc
N_Geneseq_36:T11271	+	6.00	107.23	226.14	69	Hepatitis C virus partial 5'-UTR
N_Geneseq_36:T71502	+	6.00	107.02	232.27	71	Glioblastoma U251 cell line liga
N_Geneseq_36:T59114	-	6.00	106.82	238.38	73	DNA polymerase ligand to a Therm
N_Geneseq_36:V11323	-	6.00	105.37	286.94	89	pDR-08 D-Arginine binding RNA se
N_Geneseq_36:V11321	-	6.00	105.29	289.96	90	pDR-87 D-Arginine binding RNA se
N_Geneseq_36:V11325	-	6.00	105.21	292.97	91	pDR-17 D-Arginine binding RNA se
N_Geneseq_36:X17683	-	6.00	105.13	295.98	92	Recombinant BMV RNA3 of PN10(-)
N_Geneseq_36:Q84567	-	6.00	104.52	320.00	100	HSV type 1 IE4 and IE4 pre-mess
N_Geneseq_36:T48787	+	6.00	104.10	334.94	105	Polyclonal anti-ferritin binder
N_Geneseq_36:V76263	+	6.00	104.10	337.92	106	Staphylococcus aureus contig SE
N_Geneseq_36:T39645	+	6.00	103.63	358.76	113	Oligonucleotide used to constru
N_Geneseq_36:T39636	+	6.00	103.63	358.76	113	Oligonucleotide used to constru
N_Geneseq_36:X03882	+	6.00	103.63	358.76	113	Human papillomavirus 11 L1 open
N_Geneseq_36:X03891	+	6.00	103.44	367.66	116	Human papillomavirus 11 L1 open
N_Geneseq_36:T48597	-	6.00	103.44	367.66	116	Catalytic DNA, seq09. Nucleic a
N_Geneseq_36:T48598	-	6.00	103.44	367.66	116	Catalytic DNA, seq10. Nucleic a
N_Geneseq_36:T48599	-	6.00	103.44	367.66	116	Catalytic DNA, seq11. Nucleic a
N_Geneseq_36:V13502	+	6.00	103.31	373.59	118	Hepatitis G virus nucleic acid
N_Geneseq_36:T84383	+	6.00	103.13	382.47	121	Friedreich's ataxia STM7 gene e
N_Geneseq_36:T73909	+	6.00	102.55	411.97	131	Synthetic oligonucleotide encod
N_Geneseq_36:T88358	+	6.00	102.55	411.97	131	Synthetic oligonucleotide encod
N_Geneseq_36:T24836	+	6.00	102.23	429.59	137	Human gene signature HUMGS06932
N_Geneseq_36:T11267	+	6.00	102.17	432.53	138	Hepatitis C virus partial 5'-UT
N_Geneseq_36:T11269	+	6.00	102.02	438.39	140	Hepatitis C virus partial 5'-UT
N_Geneseq_36:Q77578	+	6.00	101.86	450.10	144	Human genome fragment. (Preferr
N_Geneseq_36:N29934	+	6.00	101.52	470.54	151	Sequence of palindromic region
N_Geneseq_36:V25921	-	6.00	101.44	485.10	156	Tegenaria agrestis NPS-331 codi
N_Geneseq_36:X17679	-	6.00	100.83	514.13	166	Recombinant BMV RNA3 of PN8(-)
N_Geneseq_36:V88880	+	6.00	100.57	531.50	172	EST clone H2109. New polynucleo
N_Geneseq_36:Q79456	+	6.00	100.36	545.94	177	HCV isolate GB116 5' UTR (-244
N_Geneseq_36:Q68066	+	6.00	100.36	545.94	177	HCV isolate GB724 5' UTR (-244
N_Geneseq_36:Q68067	+	6.00	100.36	545.94	177	HCV isolate GB97 5' UTR (-244 t
N_Geneseq_36:Q68068	+	6.00	100.36	545.94	177	HCV isolate GB487 5' UTR (-244
N_Geneseq_36:Q68069	+	6.00	100.23	545.94	177	HCV isolate BE95 5' UTR (-244 t
N_Geneseq_36:Q98249	+	6.00	100.23	554.59	180	Exon 9 of rice pyruvate orthoph
N_Geneseq_36:Q43067	+	6.00	100.07	566.11	184	-255 to -62 region of 5' non-cc
N_Geneseq_36:Q43068	+	6.00	100.07	566.11	184	-255 to -62 region of 5' non-cc
N_Geneseq_36:V77510	+	6.00	100.07	566.11	184	Staphylococcus aureus contig SE
N_Geneseq_36:X33647	+	6.00	100.07	571.87	186	Hepatitis C virus partial 5'-UT
N_Geneseq_36:X11504	+	6.00	99.69	594.85	194	DNA tandem.nucleotide repeat lc
N_Geneseq_36:Q11011	+	6.00	99.32	623.49	204	Human biallelic polymorphic DNA
N_Geneseq_36:Q05051	+	6.00	98.61	683.34	225	Sequence encoding AaIT. Functic
N_Geneseq_36:T89401	+	6.00	98.57	686.19	226	Recombinant AaIt gene for insec
N_Geneseq_36:X40618	+	6.00	98.54	689.03	227	Methods for diagnosing Friedrei
N_Geneseq_36:V70455	+	6.00	98.26	714.55	236	Human secreted protein 5' EST S
N_Geneseq_36:V99109	+	6.00	98.17	723.05	239	Partial sequence of HCV subtype
N_Geneseq_36:X15319	+	6.00	98.08	731.53	242	DNA methyltransferase genomic D
N_Geneseq_36:Q60591	+	6.00	98.05	734.36	243	DNA encoding an enzyme having a
N_Geneseq_36:V86817	+	6.00	97.96	742.84	246	Human brain.Expressed Sequence
N_Geneseq_36:Q10304	-	6.00	97.87	751.31	249	EST clone BG3B6. New polynuclec
N_Geneseq_36:X40618	+	6.00	97.81	756.96	251	Human secreted protein 5' EST S
N_Geneseq_36:X12504	-	6.00	97.81	756.96	251	Human biallelic polymorphic DNA
N_Geneseq_36:X10304	-	6.00	97.81	756.96	251	Human biallelic polymorphic DNA
N_Geneseq_36:X10305	-	6.00	97.81	756.96	251	Human biallelic polymorphic DNA
N_Geneseq_36:V02955	+	6.00	97.61	776.69	258	Hepatitis GB virus (HGV) type C
N_Geneseq_36:Q29981	+	6.00	97.44	793.58	264	M2 subunit of human ribonucleot
N_Geneseq_36:V05736	+	6.00	97.41	796.39	265	Nucleotide sequence of the 5' p

Right table:

ID	Strand		Score	Value	Len	Description
N_Geneseq_36:T11816	+	6.00	97.17	821.67	274	M. kansasii ATCC 22478 (-ITG
N_Geneseq_36:V90164	+	6.00	97.12	827.28	276	EST clone DF478. New polynuc
N_Geneseq_36:V05724	+	6.00	97.01	838.49	280	Nucleotide sequence of the 3
N_Geneseq_36:X39526	-	6.00	96.91	841.29	281	Human secreted protein 5' ES
N_Geneseq_36:Q92004	+	6.00	96.91	849.69	284	Synthetic gene encoding the
N_Geneseq_36:Q11012	+	6.00	96.88	852.49	285	SP-IT gene. Functional, inse
N_Geneseq_36:Q26399	+	6.00	96.83	858.09	287	pCIB4223. Insecticidal bacul
N_Geneseq_36:X40575	+	6.00	96.44	902.76	303	Human secreted protein 5' ES
N_Geneseq_36:V87842	+	6.00	96.41	905.55	304	EST clone EY215. New polynuc
N_Geneseq_36:T19157	+	6.00	96.36	911.12	306	Human gene signature HUMGS00
N_Geneseq_36:V15619	+	6.00	96.36	913.91	307	Tomato partial invertase inh
N_Geneseq_36:V87619	+	6.00	96.34	913.91	307	EST clone EC32. New polynucl
N_Geneseq_36:Q60964	+	6.00	96.18	933.39	314	Human brain Expressed Sequen
N_Geneseq_36:Q33285	+	6.00	96.15	936.17	315	Korean hepatitis C virus CDN
N_Geneseq_36:T76274	+	6.00	96.11	941.73	317	Human neutrophil elastase (m
N_Geneseq_36:X54665	+	6.00	96.11	941.73	317	Neutrophil elastase (Medulla
N_Geneseq_36:V69002	+	6.00	96.08	944.51	318	DNA molecule encoding a poly
N_Geneseq_36:V86945	-	6.00	95.99	955.62	322	EST clone AY398. New polynuc
N_Geneseq_36:V90642	-	6.00	95.97	958.40	323	Nucleotide sequence of the
N_Geneseq_36:Q39906	+	6.00	95.88	969.50	327	Expressed Sequence Tag human
N_Geneseq_36:Q59318	+	6.00	95.88	969.50	327	Human brain Expressed Sequen
N_Geneseq_36:V78515	-	6.00	95.84	975.04	329	Staphylococcus aureus contig
N_Geneseq_36:V70856	-	6.00	95.81	977.81	330	Internal transcribed spacer
N_Geneseq_36:V11616	-	6.00	95.68	994.44	336	Homo sapiens adult brain clo
N_Geneseq_36:V20129	-	6.00	95.64	999.98	338	Probe (30) for microbial gen
N_Geneseq_36:Q78119	-	6.00	95.60	1.0e+03	340	Hepatitis C virus NS5B regio
N_Geneseq_36:Q78120	-	6.00	95.60	1.0e+03	340	Hepatitis C virus NS5B regio
N_Geneseq_36:T26389	-	6.00	95.24	1.1e+03	357	Human gene signature HUMGS08
N_Geneseq_36:V24905	+	6.00	95.22	1.1e+03	358	H. pylori secreted protein O
N_Geneseq_36:T22890	+	6.00	95.22	1.1e+03	358	Human gene signature HUMGS04
N_Geneseq_36:V24647	-	6.00	95.06	1.1e+03	366	H. pylori ORF 04gp1213orf11
N_Geneseq_36:V15597	-	6.00	94.73	1.1e+03	383	Papillomavirus major capsid
N_Geneseq_36:T84381	-	6.00	94.67	1.1e+03	386	Friedreich's ataxia STM7 gen
N_Geneseq_36:V15598	-	6.00	94.67	1.1e+03	386	Papillomavirus major capsid
N_Geneseq_36:V15599	-	6.00	94.67	1.1e+03	386	Papillomavirus major capsid
N_Geneseq_36:T19086	-	6.00	94.54	1.2e+03	393	Human gene signature HUMGS00
N_Geneseq_36:T19081	-	6.00	94.48	1.2e+03	396	Human gene signature HUMGS00
N_Geneseq_36:Q46001	-	6.00	94.45	1.2e+03	398	NPS-331. Insecticidal peptid
N_Geneseq_36:Q59672	-	6.00	94.45	1.2e+03	398	Human brain Expressed Sequen
N_Geneseq_36:Q96073	-	6.00	94.45	1.2e+03	398	Polypeptide component of spl
N_Geneseq_36:T01514	+	6.00	94.45	1.2e+03	398	Hepatitis C virus plasmid DN
N_Geneseq_36:T01513	+	6.00	94.45	1.2e+03	398	Hepatitis C virus plasmid DN
N_Geneseq_36:T47263	-	6.00	94.45	1.2e+03	398	(DSM 10103) human papillomav
N_Geneseq_36:V25920	-	6.00	94.45	1.2e+03	398	Tegenaria agrestis NPS-331 c
N_Geneseq_36:V78209	-	6.00	94.41	1.2e+03	400	Staphylococcus aureus contig
N_Geneseq_36:V78158	-	6.00	94.41	1.2e+03	400	Staphylococcus aureus contig
N_Geneseq_36:V77951	-	6.00	94.41	1.2e+03	400	Staphylococcus aureus contig
N_Geneseq_36:X40444	-	6.00	94.30	1.2e+03	406	Human secreted protein 5' ES
N_Geneseq_36:Q59951	-	6.00	94.28	1.2e+03	407	Human brain Expressed Sequen
N_Geneseq_36:T47256	-	6.00	94.23	1.2e+03	410	(DSM 10097) human papillomav
N_Geneseq_36:X20912	+	6.00	94.14	1.2e+03	415	Polynucleotide sequence from
N_Geneseq_36:T30730	+	6.00	94.07	1.2e+03	419	Mouse cryptdin 5 gene. New 1
N_Geneseq_36:V57210	+	6.00	93.98	1.2e+03	424	Clone R1.14 from mutant mult
N_Geneseq_36:V89078	+	6.00	93.95	1.2e+03	426	EST clone CB15. New polynucl
N_Geneseq_36:Q84934	+	6.00	93.92	1.3e+03	428	Human geta-globin gene. Line
N_Geneseq_36:X40584	-	6.00	93.83	1.3e+03	433	Human secreted protein 5' ES
N_Geneseq_36:X51917	+	6.00	93.78	1.3e+03	436	Human secreted protein 5' ES
N_Geneseq_36:Q70315	+	6.00	93.57	1.3e+03	449	Galectin 10 and 10SV related
N_Geneseq_36:Q70316	+	6.00	93.44	1.3e+03	457	Oxidoreducing avermectin DNA
N_Geneseq_36:T19048	+	6.00	93.44	1.3e+03	457	Oxidoreducing avermectin DNA
N_Geneseq_36:Q63396	+	6.00	93.41	1.3e+03	459	Canine genomic microsatellit
N_Geneseq_36:T65074	+	6.00	93.39	1.3e+03	460	Consensus sequence of mannur
N_Geneseq_36:X17811	+	6.00	93.39	1.3e+03	460	Microsatellite repeat sequen
N_Geneseq_36:T60776	-	6.00	93.26	1.4e+03	468	Mabinlin MBLII cDNA from Cap
N_Geneseq_36:T19047	-	6.00	93.20	1.4e+03	472	Human gene signature HUMGS00
N_Geneseq_36:T19048	+	6.00	93.20	1.4e+03	472	Human gene signature HUMGS00
N_Geneseq_36:X20812	+	6.00	93.14	1.4e+03	476	Human gene signature HUMGS00
N_Geneseq_36:Q67467	+	6.00	93.09	1.4e+03	479	Polynucleotide sequence from
N_Geneseq_36:V88503	+	6.00	93.06	1.4e+03	481	EST clone EK341. New polynuc
N_Geneseq_36:Q32446	+	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-3
N_Geneseq_36:X10305	+	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-3
N_Geneseq_36:Q32447	+	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-1
N_Geneseq_36:Q32453	+	6.00	93.03	1.4e+03	483	HCV core-envelope clone S1-3
N_Geneseq_36:X54661	+	6.00	92.94	1.4e+03	489	Neutrophil elastase (Medulla

Left column

Subject ID	Description	Length	E	% Identity		Strand
N_Geneseq_36:V89284	! EST clone CG279. New polynucleo	490	1.4e+03	92.93	6.00	+
N_Geneseq_36:V87761	! EST clone EQ218. New polynucleo	490	1.4e+03	92.93	6.00	+
N_Geneseq_36:V44834	! Clone HMOAC87R related to FcR-I	495	1.4e+03	92.86	6.00	+
N_Geneseq_36:N95402	! Sequence of new sigma-mu switch	499	1.4e+03	92.80	6.00	-
N_Geneseq_36:T83787	! DNA encoding a Staphylococcus a	501	1.5e+03	92.77	6.00	+
N_Geneseq_36:Q12239	! Clone 164/137 encoding PT-NANBH	504	1.5e+03	92.72	6.00	+
N_Geneseq_36:061601	! AK beta subunit T30A gene fragm	516	1.5e+03	92.55	6.00	-
N_Geneseq_36:061602	! AK wild type beta subunit gene	516	1.5e+03	92.55	6.00	-
N_Geneseq_36:V90747	! Nucleotide sequence of clone z4	518	1.5e+03	92.52	6.00	+
N_Geneseq_36:X30847	! Streptococcus pneumoniae genomi	521	1.5e+03	92.48	6.00	+
N_Geneseq_36:V88189	! EST clone FZ428. New polynucleo	524	1.5e+03	92.44	6.00	+
N_Geneseq_36:V75697	! Staphylococcus aureus contig SE	533	1.5e+03	92.32	6.00	+
N_Geneseq_36:T29090	! Human beta globin mutant gene f	534	1.5e+03	92.30	6.00	-
N_Geneseq_36:T74668	! Staphylococcus aureus contig SE	534	1.5e+03	92.30	6.00	+
N_Geneseq_36:T29103	! Human beta globin mutant gene f	536	1.6e+03	92.27	6.00	-
N_Geneseq_36:T29104	! Human beta globin mutant gene f	536	1.6e+03	92.27	6.00	-
N_Geneseq_36:T29089	! Human beta globin gene fragment	536	1.6e+03	92.27	6.00	-
N_Geneseq_36:T29091	! Human beta globin mutant gene f	536	1.6e+03	92.27	6.00	-
N_Geneseq_36:T29092	! Human beta globin mutant gene f	536	1.6e+03	92.09	6.00	-
N_Geneseq_36:V59519	! Human secreted protein gene 9	550	1.6e+03	92.07	6.00	+
N_Geneseq_36:V16967	! Nucleic acid sequence of the sp	551	1.6e+03	92.02	6.00	+
N_Geneseq_36:T14009	! LERK-6 coding sequence. Isolate	555	1.6e+03	92.02	6.00	+
N_Geneseq_36:T32700	! Lerk-6 coding sequence. New iso	555	1.6e+03	92.02	6.00	+
N_Geneseq_36:X32761	! Murine LERK-6 polypeptide encod	562	1.7e+03	91.93	6.00	+
N_Geneseq_36:X54633	! Human IL-4 receptor antisense c	565	1.7e+03	91.89	6.00	+
N_Geneseq_36:X30894	! Streptococcus pneumoniae genom	570	1.7e+03	91.83	6.00	+
N_Geneseq_36:V87869	! EST clone DI303. New polynucleo	583	1.7e+03	91.66	6.00	+
N_Geneseq_36:T96577	! Human bcl-w DNA. Nucleic acid e	583	1.7e+03	91.59	6.00	+
N_Geneseq_36:T39804	! Human bcl-w gene derivative. An	605	1.7e+03	91.39	6.00	+
N_Geneseq_36:X25134	! Enterococcus faecalis vanB gene	612	1.8e+03	91.31	6.00	+
N_Geneseq_36:Q69230	! Mouse p53bp2 gene. Identifying	623	1.8e+03	91.18	6.00	+
N_Geneseq_36:X13522	! Staphylococcus aureus contig SE	624	1.8e+03	91.17	6.00	+
N_Geneseq_36:X03003	! Enterococcus faecalis genome co	633	1.8e+03	91.06	6.00	+
N_Geneseq_36:Q55972	! Human IL-1ra BAC contiguous DNA	633	1.8e+03	91.06	6.00	+
N_Geneseq_36:T63512	! Murine interleukin-6 coding seq	642	1.8e+03	90.96	6.00	+
N_Geneseq_36:T89398	! Urease accessory molecule F DNA	642	1.8e+03	90.96	6.00	+
N_Geneseq_36:T89398	! Methods for diagnosing Friedrei	643	1.8e+03	90.95	6.00	+
N_Geneseq_36:T68272	! Methods for diagnosing Friedrei	644	1.8e+03	90.94	6.00	+
N_Geneseq_36:V62139	! HSV-2 strain SB5 Contig ID 67 D	646	1.8e+03	90.91	6.00	+
N_Geneseq_36:V21327	! Human I mu exon. New immunoglob	648	1.8e+03	90.89	6.00	+
N_Geneseq_36:V88885	! EST clone H288. New polynucleot	649	1.8e+03	90.88	6.00	+
N_Geneseq_36:Q53536	! Sequence of the coding region c	650	1.8e+03	90.87	6.00	+
N_Geneseq_36:Q53538	! Tobacco partial invertase inhib	654	1.9e+03	90.82	6.00	+
N_Geneseq_36:Q55969	! EST clone FZ1041. New polynucle	678	1.9e+03	90.76	6.00	+
N_Geneseq_36:T35199	! Mabinlin II cDNA encoding prepr	691	1.9e+03	90.56	6.00	+
N_Geneseq_36:Q06939	! Sequence used in identification	705	2.0e+03	90.42	6.00	+
N_Geneseq_36:Q67004	! H. pylori transporter protein C	713	2.0e+03	90.28	6.00	+
N_Geneseq_36:T94633	! EST clone FH16. New polynucleo	713	2.0e+03	90.19	6.00	+
N_Geneseq_36:T60728	! Sequence of the coding region c	717	2.0e+03	90.19	6.00	+
N_Geneseq_36:T30987	! Human dihydrofolate reductase (	722	2.0e+03	90.15	6.00	+
N_Geneseq_36:Q02295	! Mouse interleukin-12 p35 subuni	723	2.0e+03	90.10	6.00	+
N_Geneseq_36:T74633	! Probe for sequence encoding par	723	2.0e+03	90.09	6.00	+
N_Geneseq_36:Q54909	! Tumor necrosis factor receptor	728	2.0e+03	90.04	6.00	+
N_Geneseq_36:Q54680	! TNF-R1-DD ligand protein clone	730	2.0e+03	90.02	6.00	+
N_Geneseq_36:Q73598	! Rat proteosome zeta component.	730	2.1e+03	90.02	6.00	+
N_Geneseq_36:Q02295	! Poa pratensis protein allergen	730	2.1e+03	90.02	6.00	+
N_Geneseq_36:Q07004	! Gene coding for human ventricul	730	2.1e+03	90.02	6.00	+
N_Geneseq_36:T74633	! GGF gene coding segment L. Glia	730	2.1e+03	90.02	6.00	+
N_Geneseq_36:T60728	! GGF gene coding segment L. Glia	735	2.1e+03	89.97	6.00	+
N_Geneseq_36:T30987	! Huaman glial cell growth factor	736	2.1e+03	89.96	6.00	+
N_Geneseq_36:T13856	! Human glial growth factor codin	736	2.1e+03	89.96	6.00	+
N_Geneseq_36:X14358	! Human glial growth factor codin	747	2.1e+03	89.85	6.00	+
N_Geneseq_36:V69441	! H. pylori GHPO 345 gene. New is	753	2.1e+03	89.80	6.00	+
N_Geneseq_36:V69443	! Banana fruit ripening-related c	780	2.2e+03	89.54	6.00	+
N_Geneseq_36:T84068	! Banana fruit ripening-related c	786	2.2e+03	89.48	6.00	+
N_Geneseq_36:V36237	! DNA od ScFv D3M which binds to	787	2.2e+03	89.47	6.00	+
N_Geneseq_36:021968	! DNA encoding a S. aureus ATP de	787	2.2e+03	89.38	6.00	+
N_Geneseq_36:V43651	! pBCM8 SOD gene. Compsn. contg.	797	2.2e+03	89.38	6.00	+
N_Geneseq_36:V69442	! Mucin gene transcriptional regu	797	2.2e+03	89.32	6.00	+
N_Geneseq_36:V74741	! Banana fruit ripening-related c	797	2.2e+03	89.32	6.00	+
N_Geneseq_36:Q98295	! Recombinant abrin A variant wit	804	2.2e+03	89.32	6.00	+

Right column

Strand		% Identity	E	Length	Description	Subject ID
+	6.00	89.32	2.2e+03	804	Recombinant abrin A variant	N_Geneseq_36:Q98296
+	6.00	89.29	2.3e+03	807	Recombinant abrin A variant	N_Geneseq_36:T10398
+	6.00	89.27	2.3e+03	809	Canine SP5 cDNA clone, pD6k-	N_Geneseq_36:N80722
+	6.00	89.27	2.3e+03	809	Sequence of a canine SP-5 cD	N_Geneseq_36:N70716
+	6.00	89.27	2.3e+03	809	Staphylococcus aureus contig	N_Geneseq_36:V75238
+	6.00	89.27	2.3e+03	809	Banana fruit ripening-relate	N_Geneseq_36:V69440
+	6.00	89.26	2.3e+03	810	Banana fruit ripening-relate	N_Geneseq_36:V69471
+	6.00	89.25	2.3e+03	811	cDNA encoding a DAN and b57	N_Geneseq_36:V47532
+	6.00	89.18	2.3e+03	819	Photosynthetic organism carb	N_Geneseq_36:V26252
+	6.00	89.13	2.3e+03	825	Enterococcus faecalis gene E	N_Geneseq_36:X20008
+	6.00	89.10	2.3e+03	828	H. pylori cytoplasmic protei	N_Geneseq_36:X30633
+	6.00	89.02	2.3e+03	837	H. pylori cytoplasmic protei	N_Geneseq_36:X30634
+	6.00	89.02	2.4e+03	845	DNA encoding recombinant tra	N_Geneseq_36:T11323
+	6.00	88.95	2.4e+03	845	Gene regulating lipase expre	N_Geneseq_36:Q26710
+	6.00	88.95	2.4e+03	846	Human FcERI alpha-subunit an	N_Geneseq_36:Q27267
+	6.00	88.90	2.4e+03	851	Prostate cancer associated g	N_Geneseq_36:X39982
+	6.00	88.89	2.4e+03	852	Mucin MUC-1 gene promoter re	N_Geneseq_36:Q58030
+	6.00	88.84	2.4e+03	858	Human MUC1 gene 5' sequence.	N_Geneseq_36:Q79601
+	6.00	88.84	2.4e+03	858	Sequence specifically expres	N_Geneseq_36:N80471
+	6.00	88.74	2.4e+03	870	Viral infection gene SEQ ID	N_Geneseq_36:V06105
+	6.00	88.63	2.5e+03	883	Mycobacterium tuberculosis 8	N_Geneseq_36:V45248
+	6.00	88.62	2.5e+03	885	Prostate cancer associated g	N_Geneseq_36:X39977
+	6.00	88.62	2.5e+03	885	Viral infection gene SEQ ID	N_Geneseq_36:V06100
+	6.00	88.59	2.5e+03	888	cDNA encoding a novel human	N_Geneseq_36:V17184
+	6.00	88.56	2.5e+03	892	Human secreted protein gene	N_Geneseq_36:V59661
+	6.00	88.56	2.5e+03	900	Viral infection gene SEQ ID	N_Geneseq_36:V06126
+	6.00	88.37	2.5e+03	915	Sequence encoding serine pro	N_Geneseq_36:N80315
+	6.00	88.33	2.6e+03	920	TROMP1 DNA. Rare outer coat	N_Geneseq_36:Q97865
+	6.00	88.05	2.6e+03	957	Treponema pallidum rare oute	N_Geneseq_36:T76580
+	6.00	88.05	2.7e+03	957	S. marcesens Ssp-h1 autotran	N_Geneseq_36:T88154
+	6.00	88.02	2.7e+03	960	S. marcesens Ssp-h2 autotran	N_Geneseq_36:T88155
+	6.00	88.02	2.7e+03	960	Fusarium oxysporum non-surfa	N_Geneseq_36:V16105
+	6.00	87.84	2.7e+03	984	Deacetoxycephalosporin/C-syn	N_Geneseq_36:N81128
+	6.00	87.75	2.7e+03	996	Full length 3' UTR of prohi	N_Geneseq_36:T73851
+	6.00	87.72	2.8e+03	1001	Sequence encoding osteogeni	N_Geneseq_36:Q53146
+	6.00	87.70	2.8e+03	1004	Osteogenic protein 1B OP1B	N_Geneseq_36:Q72712
+	6.00	87.70	2.8e+03	1004	OP1(b) cDNA encoding OP1B f	N_Geneseq_36:T02606
+	6.00	87.70	2.8e+03	1004	Human osteogenic fusion pro	N_Geneseq_36:V15209
+	6.00	87.70	2.8e+03	1004	Osteogenic fusion protein O	N_Geneseq_36:X00234
+	6.00	87.60	2.8e+03	1018	DNA encoding a Staphylococc	N_Geneseq_36:T84101
+	6.00	87.58	2.8e+03	1020	Human prohibitin gene. Huma	N_Geneseq_36:Q50991
+	6.00	87.56	2.8e+03	1023	Encodes human IGFBP-5. DNA	N_Geneseq_36:Q28271
+	6.00	87.56	2.8e+03	1023	Insulin-like growth factor	N_Geneseq_36:T15231
+	6.00	87.56	2.8e+03	1023	TNF-R1-DD ligand protein cl	N_Geneseq_36:T94634
+	6.00	87.54	2.8e+03	1026	Meriplus giganteus galacta	N_Geneseq_36:T79497
+	6.00	87.52	2.8e+03	1029	Antibiotic resistance gene	N_Geneseq_36:V37115
+	6.00	87.49	2.8e+03	1033	DNA encoding a Staphylococc	N_Geneseq_36:T83891
+	6.00	87.49	2.8e+03	1033	Osteogenic fusion protein O	N_Geneseq_36:X14507
+	6.00	87.29	2.9e+03	1061	DNA encoding a Staphylococc	N_Geneseq_36:T35200
+	6.00	87.23	2.9e+03	1071	DNA encoding a S. pneumonia	N_Geneseq_36:T98723
+	6.00	87.12	3.0e+03	1086	Regulatory gene activating	N_Geneseq_36:T33237
+	6.00	87.11	3.0e+03	1088	cDNA encoding human mast ce	N_Geneseq_36:N90126
+	6.00	87.11	3.0e+03	1088	Nucleic acid SEQ ID NO:19 f	N_Geneseq_36:V80577
+	6.00	87.08	3.0e+03	1092	T-cell receptor alpha chain	N_Geneseq_36:Q91360
+	6.00	87.06	3.0e+03	1095	Lung cancer associated gene	N_Geneseq_36:X40185
+	6.00	87.03	3.0e+03	1100	DNA encoding a gidB polypep	N_Geneseq_36:V82081
+	6.00	86.97	3.1e+03	1109	Sequence encoding therapeut	N_Geneseq_36:Q78696
+	6.00	86.89	3.1e+03	1121	E. coli asd gene mutant. At	N_Geneseq_36:T62561
+	6.00	86.89	3.1e+03	1122	Human growth hormone secret	N_Geneseq_36:T68665
+	6.00	86.89	3.1e+03	1122	Human growth hormone secret	N_Geneseq_36:T69757
+	6.00	86.85	3.1e+03	1128	dnaN gene encoding beta sub	N_Geneseq_36:V71135
+	6.00	86.83	3.1e+03	1131	H. pylori cell envelope OMP	N_Geneseq_36:V25053
+	6.00	86.76	3.2e+03	1141	Enterococcus faecalis vanB	N_Geneseq_36:Q69235
+	6.00	86.64	3.2e+03	1142	Enterococcus faecalis genom	N_Geneseq_36:X13351
+	6.00	86.56	3.2e+03	1160	Corn threonine synthase cDN	N_Geneseq_36:X07174
+	6.00	86.56	3.2e+03	1173	Hepatitis C virus HC-J1 3'	N_Geneseq_36:Q29632
+	6.00	86.56	3.2e+03	1173	NANB hepatitis virus polynu	N_Geneseq_36:Q43893
+	6.00	86.55	3.2e+03	1174	Human FceRI alpha gene. Imm	N_Geneseq_36:Q51020
+	6.00	86.55	3.2e+03	1175	Insecticidal patatin isozym	N_Geneseq_36:Q72618
+	6.00	86.51	3.2e+03	1182	Sequence encoding bovine in	N_Geneseq_36:N60426
+	6.00	86.51	3.2e+03	1182	Staphylococcus aureus conti	N_Geneseq_36:V75110
+	6.00	86.49	3.2e+03	1185	Human secreted protein gene	N_Geneseq_36:V59783
-	6.00	86.44	3.3e+03	1193	Encodes alpha subunit of hu	N_Geneseq_36:Q04644

Left-hand listing:

Dir	Accession	Pos	Score	Conc	E-value	Description
-	N_Geneseq_36:T85615	1197	86.41	6.00	3.3e+03	Alpha subunit of human high af
+	N_Geneseq_36:V34204	1197	86.41	6.00	3.3e+03	Human secreted protein gene 51
-	N_Geneseq_36:V36343	1198	86.41	6.00	3.3e+03	cDNA encoding the alpha chain
-	N_Geneseq_36:Q14736	1199	86.40	6.00	3.3e+03	Human Fc(epsilon)RI alpha cDNA
-	N_Geneseq_36:V45211	1199	86.40	6.00	3.3e+03	Human endosulfine B gene. Nove
+	N_Geneseq_36:V90913	1200	86.40	6.00	3.3e+03	Nucleotide sequence of cluster
-	N_Geneseq_36:X01143	1200	86.40	6.00	3.3e+03	M. tuberculosis antigen clone
-	N_Geneseq_36:X01177	1200	86.40	6.00	3.3e+03	M. tuberculosis antigen clone
+	N_Geneseq_36:Q27378	1211	86.33	6.00	3.3e+03	85-C. Recombinant peptide(s) a
+	N_Geneseq_36:N60036	1215	86.31	6.00	3.3e+03	5'region of bacterial expressi
+	N_Geneseq_36:Q72693	1215	86.31	6.00	3.3e+03	Feedback inhibition release as
+	N_Geneseq_36:Q72694	1215	86.31	6.00	3.3e+03	Feedback inhibition release as
-	N_Geneseq_36:X00620	1215	86.31	6.00	3.3e+03	Human secreted protein gene 10
-	N_Geneseq_36:T97223	1217	86.29	6.00	3.3e+03	Host integration factor himD g
-	N_Geneseq_36:V65300	1224	86.25	6.00	3.3e+03	DNA encoding a S. pneumoniae F
+	N_Geneseq_36:V63008	1227	86.23	6.00	3.3e+03	D. immitis ankyrin nDIAnkl227
+	N_Geneseq_36:V63009	1227	86.23	6.00	3.3e+03	D. immitis ankyrin cDNA comple
+	N_Geneseq_36:V63006	1228	86.23	6.00	3.3e+03	D. immitis ankyrin nDIAnk1228
+	N_Geneseq_36:V63007	1228	86.23	6.00	3.3e+03	D. immitis ankyrin cDNA comple
+	N_Geneseq_36:V63310	1228	86.23	6.00	3.3e+03	Nucleotide nDIAnk1228 encoding
-	N_Geneseq_36:X04357	1239	86.16	6.00	3.4e+03	Human secreted protein gene 47
+	N_Geneseq_36:X39976	1240	86.16	6.00	3.4e+03	Prostate cancer associated gen
-	N_Geneseq_36:V13114	1245	86.13	6.00	3.4e+03	Satellite tobacco necrosis vir
+	N_Geneseq_36:V25117	1251	86.09	6.00	3.4e+03	H. Pylori cell envelope OMP OR
+	N_Geneseq_36:V59575	1255	86.07	6.00	3.4e+03	Feedback inhibition release ge
-	N_Geneseq_36:T18906	1258	86.05	6.00	3.4e+03	Delta 6-palmitoyl acyl carrier
+	N_Geneseq_36:Q48165	1263	86.02	6.00	3.4e+03	B. flavum AK gene. Gene DNA co
+	N_Geneseq_36:Q72692	1263	86.02	6.00	3.4e+03	Feedback inhibition release as
+	N_Geneseq_36:Q72695	1263	86.02	6.00	3.4e+03	Feedback inhibition release as
+	N_Geneseq_36:Q72696	1263	86.02	6.00	3.4e+03	Feedback inhibition release as
+	N_Geneseq_36:Q72697	1263	86.02	6.00	3.4e+03	DNA encoding histidyl-tRNA syn
+	N_Geneseq_36:T71310	1272	85.97	6.00	3.5e+03	Rice NADH-dependent reductase
+	N_Geneseq_36:T13486	1277	85.94	6.00	3.5e+03	Sequence encoding osteogenic f
-	N_Geneseq_36:Q53148	1277	85.94	6.00	3.5e+03	Sequence encoding osteogenic
+	N_Geneseq_36:Q72714	1277	85.94	6.00	3.5e+03	Osteogenic protein 1D OP1D fus
-	N_Geneseq_36:T02608	1277	85.94	6.00	3.5e+03	OP1(d) cDNA encoding OP1D fusi
+	N_Geneseq_36:V15211	1277	85.94	6.00	3.5e+03	Human osteogenic fusion protei
-	N_Geneseq_36:X00236	1277	85.89	6.00	3.5e+03	Osteogenic fusion protein OP1D
+	N_Geneseq_36:T77317	1287	85.89	6.00	3.5e+03	Class II EPSPS DNA for glyphos
+	N_Geneseq_36:T93793	1287	85.89	6.00	3.5e+03	Class II EPSP synthase (EPSPS)
-	N_Geneseq_36:V58018	1287	85.89	6.00	3.5e+03	Bacillus subtilis Class II EPS
-	N_Geneseq_36:X37458	1287	85.87	6.00	3.5e+03	Human secreted protein cDNA fr
-	N_Geneseq_36:V30003	1290	85.87	6.00	3.5e+03	Human CD18 cDNA. Oligonucleoti
-	N_Geneseq_36:V73325	1290	85.84	6.00	3.5e+03	Human CD18 pseudogene clone 6
+	N_Geneseq_36:Q21967	1295	85.84	6.00	3.5e+03	HindIII-NruI restriction fragm
+	N_Geneseq_36:Q49735	1296	85.80	6.00	3.6e+03	OP-1 gene. Producing a DNA seq
+	N_Geneseq_36:T39789	1302	85.80	6.00	3.6e+03	Mouse SH3P7 gene. Identifying
-	N_Geneseq_36:V61795	1303	85.79	6.00	3.6e+03	Human betac cytoplasmic domain
+	N_Geneseq_36:Q44450	1303	85.68	6.00	3.7e+03	Cysteine synthase. Novel gene
-	N_Geneseq_36:V34225	1324	85.66	6.00	3.7e+03	Human secreted protein gene 4
-	N_Geneseq_36:X22281	1327	85.66	6.00	3.7e+03	Human secreted protein gene 63
-	N_Geneseq_36:X30167	1328	85.63	6.00	3.7e+03	Human secreted protein gene 23
-	N_Geneseq_36:V34157	1332	85.63	6.00	3.7e+03	Human secreted protein gene 4
+	N_Geneseq_36:X37485	1338	85.60	6.00	3.8e+03	Streptococcus pneumoniae genom
-	N_Geneseq_36:X17775	1340	85.59	6.00	3.8e+03	HTLAR33 encoding cDNA (EST der
-	N_Geneseq_36:X30759	1348	85.55	6.00	3.8e+03	Streptococcus pneumoniae genom
+	N_Geneseq_36:V59747	1350	85.54	6.00	3.8e+03	Human secreted protein gene 89
-	N_Geneseq_36:X14148	1357	85.46	6.00	3.9e+03	H. Pylori GHPO 422 gene. New 1
+	N_Geneseq_36:X13925	1364	85.41	6.00	3.9e+03	H. Pylori GHPO 541 gene. New 1
-	N_Geneseq_36:V52382	1373	85.40	6.00	3.9e+03	Streptococcus pneumoniae genom
-	N_Geneseq_36:X04848	1373	85.39	6.00	3.9e+03	Human lysophosphatidic acid ac
-	N_Geneseq_36:V34157	1376	85.32	6.00	3.8e+03	Human secreted protein gene 4
-	N_Geneseq_36:T61118	1378	85.31	6.00	3.8e+03	DNA encoding modified OP-1. TI
-	N_Geneseq_36:Q27816	1391	85.28	6.00	3.8e+03	Adrenalin receptor subtype bet
-	N_Geneseq_36:T66442	1392	85.27	6.00	3.8e+03	Amidase gene. Nucleic acid enc
-	N_Geneseq_36:X32491	1398	85.23	6.00	3.9e+03	Seq ID NO:6 used in a method f
-	N_Geneseq_36:Q80524	1401	85.22	6.00	3.9e+03	Oxidoreducing avermectin DNA f
+	N_Geneseq_36:V59599	1407	85.19	6.00	3.9e+03	Human secreted protein gene 89
+	N_Geneseq_36:T15579	1410	85.12	6.00	3.9e+03	ICB-like apoptosis protease-2
-	N_Geneseq_36:Q44494	1416	85.11	6.00	3.9e+03	Bacillus coagulans proline. im
+	N_Geneseq_36:V24978	1430	85.07	6.00	3.9e+03	H. Pylori secreted protein ORF

Right-hand listing:

Dir	Accession	Pos	Score	Conc	E-value	Description
+	N_Geneseq_36:V24979	1440	85.07	6.00	3.9e+03	H. pylori secreted protein
+	N_Geneseq_36:V25102	1440	85.07	6.00	3.9e+03	H. Pylori secreted protein
-	N_Geneseq_36:V25111	1440	85.07	6.00	3.9e+03	H. Pylori secreted protein
-	N_Geneseq_36:Q61599	1443	85.05	6.00	3.9e+03	AK alpha subunit T279A gene
-	N_Geneseq_36:Q61600	1443	85.05	6.00	3.9e+03	AK wild type alpha subunit
-	N_Geneseq_36:T93678	1446	85.04	6.00	3.9e+03	MITL glycosidase 29G encod
-	N_Geneseq_36:V36907	1446	85.04	6.00	3.9e+03	Desulfurococcus MITL glyco
-	N_Geneseq_36:Q06188	1448	85.03	6.00	3.9e+03	Human Bone Morphogenesis Pr
-	N_Geneseq_36:Q03856	1448	85.03	6.00	3.9e+03	BMP7. Synergistic compsn. f
-	N_Geneseq_36:Q41293	1448	85.03	6.00	3.9e+03	Human BMP-7 gene. Recombina
+	N_Geneseq_36:T98605	1448	85.03	6.00	3.9e+03	DNA encoding a S. pneumonia
+	N_Geneseq_36:X24882	1448	85.03	6.00	3.9e+03	Transforming growth factor-
+	N_Geneseq_36:T31552	1456	84.98	6.00	3.9e+03	Human Ice-ced-3 homologue-1
-	N_Geneseq_36:T90555	1457	84.93	6.00	3.9e+03	Cell death protein ICH-11 e
-	N_Geneseq_36:X13315	1468	84.93	6.00	4.0e+03	Enterococcus faecalis genom
+	N_Geneseq_36:Q14857	1473	84.90	6.00	4.0e+03	Fusarium oxysporum DSM 2672
+	N_Geneseq_36:Q26407	1473	84.90	6.00	4.0e+03	Fusarium oxysporum DSM 2672
-	N_Geneseq_36:Q26382	1473	84.90	6.00	4.0e+03	Endoglucanase #2. Compact,
-	N_Geneseq_36:Q25933	1473	84.90	6.00	4.0e+03	Cellulase contained in a de
-	N_Geneseq_36:Q29935	1473	84.90	6.00	4.0e+03	Endoglucanase gene. Laundry
-	N_Geneseq_36:Q41733	1473	84.90	6.00	4.0e+03	Dye transfer inhibiting com
-	N_Geneseq_36:Q49942	1473	84.90	6.00	4.0e+03	Endoglucanase enzyme. Deter
-	N_Geneseq_36:V16103	1473	84.90	6.00	4.0e+03	Fusarium oxysporum surface-
+	N_Geneseq_36:N71068	1482	84.86	6.00	4.0e+03	Sequence encoding equine in
+	N_Geneseq_36:V60296	1486	84.84	6.00	4.0e+03	Human interleukin-1 nucleot
-	N_Geneseq_36:X00698	1487	84.83	6.00	4.0e+03	Human secreted protein gene
+	N_Geneseq_36:V56039	1488	84.83	6.00	4.0e+03	Thermostable glucose-6-phos
+	N_Geneseq_36:T89397	1490	84.82	6.00	4.0e+03	Methods for diagnosing Frie
+	N_Geneseq_36:T89397	1490	84.82	6.00	4.0e+03	Methods for diagnosing Frie
+	N_Geneseq_36:N70791	1499	84.77	6.00	4.0e+03	Sequence of ptz gene found
+	N_Geneseq_36:T36148	1499	84.77	6.00	4.0e+03	Human homologue of murine I
+	N_Geneseq_36:Q21726	1500	84.76	6.00	4.0e+03	Parainfluenza type 4A virus
+	N_Geneseq_36:T89405	1501	84.76	6.00	4.0e+03	Methods for diagnosing Frie
+	N_Geneseq_36:T89405	1501	84.76	6.00	4.0e+03	Methods for diagnosing Frie
-	N_Geneseq_36:Q53147	1505	84.74	6.00	4.0e+03	Sequence encoding osteogeni
+	N_Geneseq_36:Q72713	1505	84.74	6.00	4.0e+03	Osteogenic protein 1C OPIC
-	N_Geneseq_36:T02607	1505	84.74	6.00	4.0e+03	OP1(c) cDNA encoding OP1C
+	N_Geneseq_36:V15210	1505	84.74	6.00	4.0e+03	Human osteogenic fusion pro
-	N_Geneseq_36:X00235	1506	84.74	6.00	4.0e+03	Osteogenic fusion protein O
+	N_Geneseq_36:Q29143	1506	84.74	6.00	4.0e+03	DM beta cDNA. Identifying m
-	N_Geneseq_36:T39628	1506	84.74	6.00	4.0e+03	HPV6/11 hybrid DNA encoding
-	N_Geneseq_36:V16721	1506	84.74	6.00	4.0e+03	Nucleic acid encoding rat N
-	N_Geneseq_36:V33585	1506	84.74	6.00	4.1e+03	Human papillomavirus 6/11 h
-	N_Geneseq_36:N81125	1509	84.72	6.00	4.1e+03	Deacetoxycephalosporin-C-sy
+	N_Geneseq_36:Q53470	1509	84.72	6.00	4.1e+03	PTK gene B1. Expression of
+	N_Geneseq_36:V59788	1517	84.69	6.00	4.1e+03	Nucleic acid encoding human
+	N_Geneseq_36:V16720	1518	84.68	6.00	4.1e+03	N-Wiskott-Aldrich syndrome
+	N_Geneseq_36:V20072	1518	84.68	6.00	4.1e+03	M. vaccae pota homologue en
+	N_Geneseq_36:V34595	1518	84.68	6.00	4.1e+03	DNA of genomic clone encodi
-	N_Geneseq_36:Q02852	1519	84.66	6.00	4.1e+03	Human secreted protein gene
-	N_Geneseq_36:X22273	1523	84.65	6.00	4.1e+03	Human papillomavirus type 1
-	N_Geneseq_36:Q12773	1524	84.64	6.00	4.1e+03	DNA encoding lysophosphatid
-	N_Geneseq_36:T85928	1527	84.63	6.00	4.1e+03	Human p69 cDNA. Antigen com
-	N_Geneseq_36:T05577	1528	84.58	6.00	4.1e+03	Mycobacterium tuberculosis
-	N_Geneseq_36:V44370	1539	84.58	6.00	4.1e+03	M. tuberculosis immunogenic
-	N_Geneseq_36:V64478	1539	84.57	6.00	4.2e+03	Streptococcus pneumoniae pr
+	N_Geneseq_36:V72588	1542	84.53	6.00	4.2e+03	H. Pylori GHPO 1102 gene.
+	N_Geneseq_36:X14075	1550	84.52	6.00	4.2e+03	Borrelia burgdorferi polynu
-	N_Geneseq_36:X20313	1552	84.46	6.00	4.2e+03	cDNA encoding transforming
-	N_Geneseq_36:T47652	1565	84.45	6.00	4.3e+03	Human polymorphic epithelia
-	N_Geneseq_36:Q12017	1567	84.33	6.00	4.3e+03	Rat acetylglucosaminyl tran
-	N_Geneseq_36:T29199	1593	84.33	6.00	4.3e+03	Human N-acetylglucosaminyl
-	N_Geneseq_36:T77094	1593	84.29	6.00	4.3e+03	Recombinant papilloma virus
-	N_Geneseq_36:T05834	1599	84.29	6.00	4.3e+03	Human mitogen-activated pro
-	N_Geneseq_36:T43204	1602	84.29	6.00	4.3e+03	Human mitogen activated pro
-	N_Geneseq_36:X07066	1602	84.28	6.00	4.3e+03	Acyl-ACP thioesterase. Modi
-	N_Geneseq_36:Q26528	1603	84.25	6.00	4.3e+03	IGFBP6 DNA. Insulin-like gr
-	N_Geneseq_36:Q26814	1611	84.24	6.00	4.3e+03	DNA sequence of glycosyltra
-	N_Geneseq_36:X26196	1613	84.23	6.00	4.3e+03	C6 insertion plasmid pC6L.
+	N_Geneseq_36:Q67875	1614	84.23	6.00	4.3e+03	ALVAC C6 flanking arms. Nuc
+	N_Geneseq_36:T01460	1615	84.23	6.00	4.3e+03	ALVAC C6 flanking arm used
+	N_Geneseq_36:T04704	1615	84.23	6.00	4.3e+03	Canarypox virus ALVAC pC6L
+	N_Geneseq_36:V24972	1615	84.23	6.00	4.3e+03	

Left column

ID		E-value	Score	Str		Description	Length
N_Geneseq_36:V27175	6.00	4.3e+03	84.18	-	!	G-protein coupled receptor, 1c	1626
N_Geneseq_36:Q54496	6.00	4.4e+03	84.17	-	!	Alcaligenes bronchisepticus ar	1629
N_Geneseq_36:N40071	6.00	4.4e+03	84.14	-	!	Sequence coding for Herpes sim	1635
N_Geneseq_36:T51325	6.00	4.4e+03	84.14	-	!	HSV glycoprotein D-2 DNA. Vari	1635
N_Geneseq_36:V31670	6.00	4.4e+03	84.14	+	!	Herpes simplex virus type 2 gl	1637
N_Geneseq_36:Q10429	6.00	4.4e+03	84.13	+	!	Male hormone-dependent gene. N	1639
N_Geneseq_36:V80562	6.00	4.4e+03	84.12	+	!	Human lysophosphatidic acid ac	1642
N_Geneseq_36:Q14835	6.00	4.4e+03	84.11	+	!	Human serotonin receptor-encod	1643
N_Geneseq_36:Q61596	6.00	4.4e+03	84.11	+	!	AK wild type alpha subunit gen	1643
N_Geneseq_36:Q61597	6.00	4.4e+03	84.10	+	!	AK beta subunit T30A gene. The	1643
N_Geneseq_36:Q61598	6.00	4.4e+03	84.10	+	!	AK wild type beta subunit gene	1643
N_Geneseq_36:Q61595	6.00	4.4e+03	84.10	+	!	AK alpha subunit T279A gene. T	1643
N_Geneseq_36:Q84971	6.00	4.4e+03	84.10	-	!	C. glutamicum gene. Variant of	1643
N_Geneseq_36:T03326	6.00	4.4e+03	84.10	-	!	C. glutamicum gene. Variant of	1643
N_Geneseq_36:T49287	6.00	4.4e+03	84.10	-	!	Fragment of the B. lactofermen	1643
N_Geneseq_36:T49288	6.00	4.4e+03	84.10	-	!	DNA encoding B. lactofermentum	1643
N_Geneseq_36:T49289	6.00	4.4e+03	84.10	-	!	DNA encoding B. lactofermentum	1643
N_Geneseq_36:T84785	6.00	4.4e+03	84.10	-	!	Brevibacterium lactofermentum	1643
N_Geneseq_36:V15781	6.00	4.4e+03	84.10	-	!	B. lactofermentum aspartokinas	1643
N_Geneseq_36:V15782	6.00	4.4e+03	84.10	-	!	B. lactofermentum aspartokinas	1643
N_Geneseq_36:V40254	6.00	4.4e+03	84.10	-	!	Brevibacterium lactofermentum	1643
N_Geneseq_36:V40255	6.00	4.4e+03	84.10	-	!	Brevibacterium lactofermentum	1643
N_Geneseq_36:V41300	6.00	4.4e+03	84.10	-	!	Aspartokinase lysC gene. New	1643
N_Geneseq_36:V41316	6.00	4.4e+03	84.10	-	!	Mutant aspartokinase lysC gene	1643
N_Geneseq_36:Q65519	6.00	4.4e+03	84.07	-	!	Human insulin-like growth fact	1650
N_Geneseq_36:T72216	6.00	4.6e+03	84.07	-	!	Grapevine leafroll virus hsp-7	1650
N_Geneseq_36:Q80218	6.00	4.4e+03	84.07	-	!	Human NDF-alpha2b clone 17 DNA	1651
N_Geneseq_36:Q66990	6.00	4.4e+03	84.05	-	!	DF3 enhancer. Isolated DF3 enh	1656
N_Geneseq_36:V77543	6.00	4.4e+03	84.05	-	!	Staphylococcus aureus contig S	1656
N_Geneseq_36:T45050	6.00	4.4e+03	84.00	-	!	Beta-globin gene for retrovira	1666
N_Geneseq_36:N70926	6.00	4.5e+03	83.94	-	!	Plasmid pTUB616 encoding heat	1680
N_Geneseq_36:V34545	6.00	4.5e+03	83.94	-	!	Saccharomyces cerevisiae RNA 3	1680
N_Geneseq_36:V54134	6.00	4.5e+03	83.92	-	!	Sequence encoding mouse adipoc	1685
N_Geneseq_36:T44455	6.00	4.5e+03	83.92	-	!	cDNA encoding murine p154. Man	1685
N_Geneseq_36:T08083	6.00	4.5e+03	83.85	+	!	Alpha-1,3/4-fucosidase gene. A	1689
N_Geneseq_36:N70538	6.00	4.6e+03	83.79	+	!	Sequence encoding alpha-amylas	1700
N_Geneseq_36:T04588	6.00	4.6e+03	83.78	-	!	Rat Vhh-1 cDNA. Nucleic acid	1715
N_Geneseq_36:Q92299	6.00	4.6e+03	83.78	-	!	Mango class II thioesterase MA	1717
N_Geneseq_36:Q15233	6.00	4.6e+03	83.77	-	!	Class II acyl-ACP thioesterase	1717
N_Geneseq_36:N70920	6.00	4.6e+03	83.77	+	!	Plasmid pTUB613 encoding heat	1719
N_Geneseq_36:Q66988	6.00	4.6e+03	83.75	+	!	5' flanking region oof the DF3	1725
N_Geneseq_36:N90523	6.00	4.6e+03	83.74	+	!	Herpes Simplex virus-2 gD gene	1726
N_Geneseq_36:N50488	6.00	4.6e+03	83.73	+	!	Sequence of herpes simplex vir	1730
N_Geneseq_36:X23754	6.00	4.6e+03	83.71	-	!	Rice anthranilate synthase fir	1734
N_Geneseq_36:X23748	6.00	4.6e+03	83.71	-	!	Rice anthranilate synthase fir	1734
N_Geneseq_36:T85939	6.00	4.6e+03	83.67	+	!	Shigella flexneri IpaB gene. S	1743
N_Geneseq_36:r92839	6.00	4.7e+03	83.61	-	!	Flea saliva protein fspMF) cD	1758
N_Geneseq_36:T73387	6.00	4.7e+03	83.61	-	!	Flea saliva protein nucleic ac	1758
N_Geneseq_36:Q12771	6.00	4.7e+03	83.59	+	!	Human polymorphic epithelial	1763
N_Geneseq_36:Q79143	6.00	4.7e+03	83.58	+	!	Hepatitis C virus gene SRQ37-5	1765
N_Geneseq_36:V59594	6.00	4.7e+03	83.54	+	!	Human secreted protein gene 84	1774
N_Geneseq_36:V47531	6.00	4.7e+03	83.52	+	!	cDNA encoding a DAN and b57 pr	1781
N_Geneseq_36:Q42191	6.00	4.8e+03	83.50	-	!	PM-1 antigen. Neuro-endocrine	1785
N_Geneseq_36:T04154	6.00	4.8e+03	83.47	-	!	BCG DapB gene. Gene involved	1791
N_Geneseq_36:T03250	6.00	4.8e+03	83.46	-	!	Enterobacter sp. S262 sucrose	1794
N_Geneseq_36:V22703	6.00	4.8e+03	83.45	-	!	Human recombinase hsREC2 cDNA.	1797
N_Geneseq_36:Q05999	6.00	4.8e+03	83.44	-	!	Sequence encoding foetal human	1800
N_Geneseq_36:T90412	6.00	4.8e+03	83.44	-	!	Human glutamate transporter pr	1800
N_Geneseq_36:V31149	6.00	4.8e+03	83.44	-	!	Human excitatory amino acid tr	1800
N_Geneseq_36:V61477	6.00	4.8e+03	83.44	-	!	Human secreted protein bd164_7	1800
N_Geneseq_36:V69231	6.00	4.8e+03	83.44	-	!	Human excitatory amino acid tr	1800
N_Geneseq_36:T03249	6.00	4.8e+03	83.43	+	!	Protaminobacter rubrum sucrose	1803
N_Geneseq_36:Q04673	6.00	4.8e+03	83.42	+	!	Sul gene containg fragment (1.	1804
N_Geneseq_36:T15008	6.00	4.8e+03	83.40	+	!	Mouse Elf-1 cDNA. Murine and	1809
N_Geneseq_36:V42926	6.00	4.8e+03	83.38	+	!	cDNA encoding a mammalian Elf-	1809
N_Geneseq_36:T90043	6.00	4.8e+03	83.38	-	!	Bacillus stearothermophilus al	1814
N_Geneseq_36:V95033	6.00	4.8e+03	83.38	-	!	Bacillus stearothermophilus al	1814
N_Geneseq_36:Q88068	6.00	4.8e+03	83.38	-	!	Bacillus stearothermophilus Te	1814
N_Geneseq_36:V02473	6.00	4.8e+03	83.36	-	!	cDNA sequence from mRNA of glu	1820
N_Geneseq_36:Q04731	6.00	4.8e+03	83.35	-	!	Sequence encoding mammalian os	1822
N_Geneseq_36:Q11803	6.00	4.8e+03	83.35	-	!	Human osteogenic protein hOP1	1822
N_Geneseq_36:Q24518	6.00	4.8e+03	83.35	-	!	Human osteogenic protein hOP1	1822
N_Geneseq_36:Q28735	6.00	4.8e+03	83.35	-	!	Human osteogenic protein hOP1	1822

Right column

ID		E-value	Score	Str		Description	Length
N_Geneseq_36:V38944	6.00	4.8e+03	83.35	-	!	Human osteogenic protein 1	1822
N_Geneseq_36:Q38733	6.00	4.8e+03	83.35	-	!	Human OP-1. Morphogen-induc	1822
N_Geneseq_36:Q38857	6.00	4.8e+03	83.35	-	!	Morphogen hOP1 coding seque	1822
N_Geneseq_36:Q38141	6.00	4.8e+03	83.35	-	!	Sequence encoding human ost	1822
N_Geneseq_36:Q56198	6.00	4.8e+03	83.35	-	!	hOP1 cDNA. Morphogenic prot	1822
N_Geneseq_36:Q57916	6.00	4.8e+03	83.35	-	!	Human osteogenic protein hO	1822
N_Geneseq_36:T07311	6.00	4.8e+03	83.35	-	!	Human OP-1. Morphogen-induc	1822
N_Geneseq_36:Q45116	6.00	4.8e+03	83.35	-	!	Human OP-1. Maintaining int	1822
N_Geneseq_36:Q65391	6.00	4.8e+03	83.35	-	!	Osteogenic protein hOP1-PP.	1822
N_Geneseq_36:Q45162	6.00	4.8e+03	83.35	-	!	Human OP-1. Use of morphoge	1822
N_Geneseq_36:Q54755	6.00	4.8e+03	83.35	-	!	Osteogenic protein OP1. Bon	1822
N_Geneseq_36:Q71425	6.00	4.8e+03	83.35	-	!	Osteogenic protein OP1 cDNA	1822
N_Geneseq_36:Q72703	6.00	4.8e+03	83.35	-	!	hOP1 human osteogenic prote	1822
N_Geneseq_36:T06038	6.00	4.8e+03	83.35	-	!	Human osteogenic protein 1	1822
N_Geneseq_36:T02597	6.00	4.8e+03	83.35	-	!	hOP-1 cDNA. Antibodies with	1822
N_Geneseq_36:T06771	6.00	4.8e+03	83.35	-	!	Human osteogenic protein OP	1822
N_Geneseq_36:T33441	6.00	4.8e+03	83.35	-	!	Human osteogenic protein OP	1822
N_Geneseq_36:T73207	6.00	4.8e+03	83.35	-	!	Human hippocampal osteogeni	1822
N_Geneseq_36:T87879	6.00	4.8e+03	83.35	-	!	Human osteogenic protein 1	1822
N_Geneseq_36:T97877	6.00	4.8e+03	83.35	-	!	cDNA encoding human osteoge	1822
N_Geneseq_36:T93810	6.00	4.8e+03	83.35	-	!	Human osteogenic protein-1	1822
N_Geneseq_36:Q10345	6.00	4.8e+03	83.35	-	!	Human OP-1 cDNA. Treatment	1822
N_Geneseq_36:V15204	6.00	4.8e+03	83.35	-	!	Human osteogenic protein OP	1822
N_Geneseq_36:V19533	6.00	4.8e+03	83.35	-	!	Human osteogenic protein (h	1822
N_Geneseq_36:V41572	6.00	4.8e+03	83.35	-	!	Human osteogenic protein 1	1822
N_Geneseq_36:V60227	6.00	4.8e+03	83.35	-	!	Morphogenically active oest	1822
N_Geneseq_36:V65199	6.00	4.8e+03	83.35	-	!	Nucleotide sequence encodin	1822
N_Geneseq_36:V80656	6.00	4.8e+03	83.35	-	!	Human osteogenic protein cD	1822
N_Geneseq_36:V80735	6.00	4.8e+03	83.35	-	!	Human osteogenic protein OP	1822
N_Geneseq_36:Q56231	6.00	4.8e+03	83.35	-	!	Human osteogenic protein 1	1822
N_Geneseq_36:Q56231	6.00	4.8e+03	83.35	-	!	Human osteogenic protein OP	1822
N_Geneseq_36:Q87837	6.00	4.8e+03	83.35	-	!	hOP1 cDNA. Use morphogens t	1822
N_Geneseq_36:Q14627	6.00	4.8e+03	83.33	-	!	Human calcium channel subun	1823
N_Geneseq_36:T34367	6.00	4.9e+03	83.33	-	!	Tomato phosphate transporte	1826
N_Geneseq_36:T33630	6.00	4.9e+03	83.32	+	!	Human Glioblastoma cell cDN	1829
N_Geneseq_36:V53631	6.00	4.9e+03	83.32	+	!	Plasmid pJC265 (ATCC 68598)	1829
N_Geneseq_36:T34420	6.00	4.9e+03	83.27	+	!	Rat p65 oncofoetal protein	1836
N_Geneseq_36:V12887	6.00	5.0e+03	83.16	-	!	Human 7-transmembrane recep	1841
N_Geneseq_36:Q90623	6.00	5.0e+03	83.16	-	!	Pyranose oxidase encoding s	1869
N_Geneseq_36:T34442	6.00	5.0e+03	83.15	-	!	Thermococcus amidase gene.	1869
N_Geneseq_36:Q24517	6.00	5.0e+03	83.15	-	!	Murine osteogenic protein (	1872
N_Geneseq_36:Q28736	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein OP	1873
N_Geneseq_36:Q38945	6.00	5.0e+03	83.15	-	!	Murine osteogenic protein m	1873
N_Geneseq_36:Q32584	6.00	5.0e+03	83.15	-	!	Murine osteogenic protein m	1873
N_Geneseq_36:Q38858	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein 1	1873
N_Geneseq_36:Q53153	6.00	5.0e+03	83.15	-	!	Murine pro-OP-1. Morphogen-	1873
N_Geneseq_36:Q51199	6.00	5.0e+03	83.15	-	!	Morphogen mOP1 coding seque	1873
N_Geneseq_36:Q56232	6.00	5.0e+03	83.15	-	!	Sequence encoding murine os	1873
N_Geneseq_36:Q58051	6.00	5.0e+03	83.15	-	!	Murine OP-1. Morphogen-indu	1873
N_Geneseq_36:Q67312	6.00	5.0e+03	83.15	-	!	mOP1 cDNA. Morphogenic prot	1873
N_Geneseq_36:Q45117	6.00	5.0e+03	83.15	-	!	mOP1 cDNA. Use morphogens t	1873
N_Geneseq_36:Q45392	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein mO	1873
N_Geneseq_36:Q45163	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein OP	1873
N_Geneseq_36:X00241	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein 1	1873
N_Geneseq_36:Q72704	6.00	5.0e+03	83.15	-	!	Murine OP-1. Maintaining in	1873
N_Geneseq_36:Q80220	6.00	5.0e+03	83.15	-	!	Osteogenic protein mOP1-PP.	1873
N_Geneseq_36:T02598	6.00	5.0e+03	83.15	-	!	Murine OP-1. Use of morphog	1873
N_Geneseq_36:T87879	6.00	5.0e+03	83.15	-	!	mOP1-PP prepro form mouse o	1873
N_Geneseq_36:V10346	6.00	5.0e+03	83.15	-	!	Human NDF-beta1 clone P13	1873
N_Geneseq_36:V10347	6.00	5.0e+03	83.15	-	!	mOP-1 cDNA. Antibodies with	1873
N_Geneseq_36:V19534	6.00	5.0e+03	83.15	-	!	cDNA encoding mouse osteoge	1873
N_Geneseq_36:V32584	6.00	5.0e+03	83.15	-	!	Mouse OP-1 cDNA. Treatment	1873
N_Geneseq_36:X00241	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein OP	1873
N_Geneseq_36:Q33033	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein OP	1873
N_Geneseq_36:Q33017	6.00	5.0e+03	83.15	-	!	Mouse osteogenic protein 1	1873
N_Geneseq_36:T61117	6.00	5.0e+03	83.14	-	!	Murine osteogenic protein M	1875
N_Geneseq_36:T03246	6.00	5.0e+03	83.14	+	!	Dirofilaria immitis transgl	1875
N_Geneseq_36:V52352	6.00	5.0e+03	83.08	+	!	Dirofilaria immitis transgl	1878
N_Geneseq_36:V53032	6.00	5.0e+03	83.04	+	!	DNA encoding OP-1. Transfor	1890
N_Geneseq_36:Q11055	6.00	5.0e+03	82.97	-	!	Protaminobacter rubrum sucr	1900
N_Geneseq_36:T31931	6.00	5.1e+03	82.96	-	!	Streptococcus pneumoniae ge	1906
N_Geneseq_36:V59546	6.00	5.1e+03	82.96	-	!	CD30 ligand gene used in th	1919
N_Geneseq_36:	6.00	5.1e+03	82.96	-	!	Clone 19 encoding murine IL	1922
N_Geneseq_36:	6.00	5.1e+03	82.95	-	!	Retinoid X receptor interac	1922
N_Geneseq_36:	6.00	5.1e+03	82.95	+	!	Human secreted protein gene	1924

ID						Description
N_Geneseq_36:X08426	-	6.00	82.93	5.1e+03	1929	Acidic leucine aminopeptidase
N_Geneseq_36:Q04472	+	6.00	82.85	5.2e+03	1950	Human papilloma virus 52 clone
N_Geneseq_36:Q56791	+	6.00	82.78	5.2e+03	1971	cDNA encoding receptor for C-t
N_Geneseq_36:X08428	-	6.00	82.78	5.2e+03	1971	Rat beta-amyloid precursor pro
N_Geneseq_36:T67025	-	6.00	82.73	5.2e+03	1971	Acidic leucine aminopeptidase
N_Geneseq_36:Q28118	-	6.00	82.73	5.3e+03	1983	Human norepinepherine transpor
N_Geneseq_36:X33115	-	6.00	82.72	5.3e+03	1983	Human PRL-1 phosphatase encodi
N_Geneseq_36:V69212	+	6.00	82.68	5.3e+03	1986	C. acidovorans polyurethane e
N_Geneseq_36:T90830	+	6.00	82.67	5.3e+03	1996	cDNA encoding glucoanylase P.
N_Geneseq_36:Q05549	+	6.00	82.67	5.3e+03	1999	Sequence encoding Newcastle di
N_Geneseq_36:N91000	+	6.00	82.67	5.3e+03	2000	Sequence encoding Newcastle di
N_Geneseq_36:T88842	+	6.00	82.63	5.3e+03	2000	Mycobacterium tuberculosis sig
N_Geneseq_36:X11129	+	6.00	82.61	5.3e+03	2001	Enterococcus faecalis genome
N_Geneseq_36:T88823	+	6.00	82.60	5.3e+03	2012	Cytosolic phospholipase A2/B c
N_Geneseq_36:N91804	-	6.00	82.59	5.3e+03	2017	Human papilloma virus type 33
N_Geneseq_36:N90641	+	6.00	82.56	5.4e+03	2019	Staphylococcus aureus mecB ORF
N_Geneseq_36:V16292	-	6.00	82.52	5.4e+03	2020	Human follistatin (hFS) gene c
N_Geneseq_36:Q31542	-	6.00	82.52	5.4e+03	2023	Tomato phosphate transporter 1
N_Geneseq_36:X01163	-	6.00	82.47	5.4e+03	2029	Human heregulin beta-1 in clon
N_Geneseq_36:X01129	+	6.00	82.45	5.4e+03	2041	M. tuberculosis antigen clone
N_Geneseq_36:Q22216	+	6.00	82.45	5.4e+03	2041	M. tuberculosis antigen clone
N_Geneseq_36:T59296	+	6.00	82.43	5.4e+03	2054	Squalene synthase. Nucleic aci
N_Geneseq_36:Q40093	+	6.00	82.40	5.4e+03	2054	S. cerevisiae squalene synthet
N_Geneseq_36:N90340	+	6.00	82.38	5.4e+03	2060	CPH1 gene. Isolated fungal dim
N_Geneseq_36:Q41746	+	6.00	82.37	5.5e+03	2061	Sequence encoding a receptor F
N_Geneseq_36:Q03288	-	6.00	82.36	5.5e+03	2066	IL-6 receptor coding sequence.
N_Geneseq_36:Q90719	+	6.00	82.36	5.5e+03	2076	Gene encoding yeast NADPH-cyto
N_Geneseq_36:Q90721	+	6.00	82.28	5.5e+03	2081	B. burgdorferi strain K48 anti
N_Geneseq_36:Q88694	+	6.00	82.27	5.6e+03	2081	B. burgdorferi strain pTrob an
N_Geneseq_36:Q81504	-	6.00	82.27	5.6e+03	2084	HindIII fragment pNF2009 (HA9)
N_Geneseq_36:N90847	+	6.00	82.26	5.6e+03	2085	OPDE 45 kDa subunit gene. New
N_Geneseq_36:Q67458	-	6.00	82.25	5.6e+03	2087	DNA contg. region encoding BSF
N_Geneseq_36:T68825	-	6.00	82.22	5.6e+03	2088	Codes tau-protein kinase I (TP
N_Geneseq_36:T68825	-	6.00	82.21	5.6e+03	2088	Cytosolic phospholipase A2/B c
N_Geneseq_36:Q05990	+	6.00	82.21	5.6e+03	2109	PstI-XhoI fragment of plasmid
N_Geneseq_36:T68826	-	6.00	82.19	5.7e+03	2112	Cytosolic phospholipase A2/B c
N_Geneseq_36:Q96055	-	6.00	82.13	5.7e+03	2112	Sequence encoding melanoma ass
N_Geneseq_36:Q12242	-	6.00	82.09	5.7e+03	2115	Encodes PT-NANBH viral structu
N_Geneseq_36:X13605	-	6.00	82.08	5.7e+03	2116	Enterococcus faecalis genome c
N_Geneseq_36:T03760	-	6.00	82.06	5.7e+03	2119	Kidney injury associated molec
N_Geneseq_36:T06726	-	6.00	82.01	5.7e+03	2127	Melanoma-specific immunogen, F
N_Geneseq_36:V01587	-	6.00	82.21	5.8e+03	2131	pMEL17 cDNA. Melanoma-specific
N_Geneseq_36:V15220	-	6.00	82.13	5.8e+03	2136	Human glycine transporter 1a e
N_Geneseq_36:T58558	-	6.00	82.09	5.8e+03	2154	Bacillus thuringiensis 158C2a
N_Geneseq_36:X04731	-	6.00	82.08	5.8e+03	2164	Human breast cancer cell glyco
N_Geneseq_36:T31553	-	6.00	82.06	5.8e+03	2164	Full cDNA sequence of the gp30
N_Geneseq_36:T02716	-	6.00	82.01	5.8e+03	2167	Human Ice-ced-3 homologue-1S
N_Geneseq_36:Q79972	-	6.00	82.00	5.9e+03	2174	MART-1 melanoma antigen cDNA25
N_Geneseq_36:X20539	-	6.00	82.00	5.9e+03	2190	Human interleukin-1 beta conve
N_Geneseq_36:V43016	+	6.00	81.99	5.9e+03	2192	Polynucleotide sequence from 6
N_Geneseq_36:X04732	-	6.00	81.99	5.9e+03	2193	Xenopus orphan receptor 6 cDNA
N_Geneseq_36:V68494	-	6.00	81.98	5.9e+03	2196	Genomic form of lambda-5. Nucl
N_Geneseq_36:T36499	-	6.00	81.98	5.9e+03	2199	Human gelsolin gene coding reg
N_Geneseq_36:Q06982	-	6.00	81.98	5.9e+03	2199	Sequence encoding heat resista
N_Geneseq_36:T16542	-	6.00	81.97	5.9e+03	2199	M.gallisepticum DNA sequence I
N_Geneseq_36:Q03543	-	6.00	81.92	5.9e+03	2201	Human betaine/GABA transporter
N_Geneseq_36:T84445	-	6.00	81.91	5.9e+03	2202	Canola cold responsive gene BN
N_Geneseq_36:N90389	-	6.00	81.86	5.9e+03	2217	Human breast cancer cell glyco
N_Geneseq_36:Q57973	+	6.00	81.83	5.9e+03	2218	Streptococcus pneumoniae polyp
N_Geneseq_36:T42784	-	6.00	81.83	5.9e+03	2235	Full cDNA sequence of the gp30
N_Geneseq_36:Q03998	+	6.00	81.82	5.9e+03	2245	Homo sapiens Gly T1d glycine t
N_Geneseq_36:Q99002	+	6.00	81.80	5.9e+03	2245	Hexokinase exon II promoter fr
N_Geneseq_36:Q99002	+	6.00	81.79	5.9e+03	2247	Human OVCA1 tumour suppressor
N_Geneseq_36:T08098	-	6.00	81.72	6.0e+03	2253	Human betaine/GABA transporter
N_Geneseq_36:T08098	+	6.00	81.72	6.0e+03	2256	Human betaine/GABA transporter
N_Geneseq_36:T43382	-	6.00	81.66	6.0e+03	2278	DNA encoding polypeptide with
					2285	Chromosomal gene fragment of H
					2297	Human STCH chaperone protein c

ID						Description
N_Geneseq_36:T47338	+	6.00	81.64	6.0e+03	2303	Murine developmentally-regu
N_Geneseq_36:T96034	-	6.00	81.64	6.0e+03	2303	Rat kidney injury related m
N_Geneseq_36:X18506	+	6.00	81.64	6.0e+03	2303	Murine del-1 encoding cDNA.
N_Geneseq_36:Q39149	-	6.00	81.63	6.0e+03	2305	Annexin XI gene in clone pC
N_Geneseq_36:X18507	-	6.00	81.62	6.0e+03	2308	Murine del-1 truncated mino
N_Geneseq_36:T47339	-	6.00	81.62	6.0e+03	2308	Truncated murine del-1 enco
N_Geneseq_36:X18507	+	6.00	81.61	6.0e+03	2311	Annexin XI gene in clone pC
N_Geneseq_36:Q39150	-	6.00	81.61	6.1e+03	2311	Mouse inward rectifier K+ c
N_Geneseq_36:Q82904	+	6.00	81.59	6.1e+03	2319	Membrane-bound aldehyde deh
N_Geneseq_36:N90816	+	6.00	81.58	6.1e+03	2323	Mouse developmental kinase
N_Geneseq_36:T32962	+	6.00	81.56	6.1e+03	2327	H. pylori GHPO 961 gene. Ne
N_Geneseq_36:X14459	+	6.00	81.54	6.1e+03	2335	Human proNDF-alpha2b DNA. N
N_Geneseq_36:Q80216	+	6.00	81.47	6.2e+03	2356	Rat NDF clone 38 DNA. New r
N_Geneseq_36:Q80229	+	6.00	81.46	6.2e+03	2361	Burkholderia cepacia insert
N_Geneseq_36:T84986	+	6.00	81.46	6.2e+03	2361	Human secreted protein gene
N_Geneseq_36:X04316	+	6.00	81.46	6.2e+03	2362	Staphylococcus epidermidis
N_Geneseq_36:Q55137	+	6.00	81.43	6.2e+03	2369	Mycoplasma gallisepticum an
N_Geneseq_36:Q77863	-	6.00	81.43	6.2e+03	2371	cDNA encoding a human chron
N_Geneseq_36:V63470	-	6.00	81.42	6.2e+03	2373	N. meningitidis serotype C
N_Geneseq_36:T26997	+	6.00	81.41	6.2e+03	2376	N. meningitidis serotype A
N_Geneseq_36:T26998	+	6.00	81.41	6.2e+03	2376	N. meningitidis serotype A
N_Geneseq_36:V72590	+	6.00	81.41	6.2e+03	2379	Mouse neurotrypsin encoding
N_Geneseq_36:T26999	+	6.00	81.40	6.2e+03	2379	N. meningitidis serotype B
N_Geneseq_36:T67024	+	6.00	81.37	6.2e+03	2390	Rat beta-amyloid precursor
N_Geneseq_36:X17774	-	6.00	81.36	6.2e+03	2394	YAK-1 related serine/threon
N_Geneseq_36:Q06622	+	6.00	81.34	6.3e+03	2400	Asparagine synthetase AS1 g
N_Geneseq_36:Q04305	+	6.00	81.25	6.3e+03	2430	Coding region of murine IL-
N_Geneseq_36:Q80232	-	6.00	81.25	6.3e+03	2430	Rat NDF clone 42A DNA. New
N_Geneseq_36:T61550	-	6.00	81.25	6.3e+03	2430	Murine interleukin-4 recept
N_Geneseq_36:V17655	-	6.00	81.25	6.3e+03	2430	Mouse interleukin-4 recepto
N_Geneseq_36:V38332	-	6.00	81.25	6.3e+03	2430	Mouse interleukin-4 recepto
N_Geneseq_36:V08858	-	6.00	81.25	6.3e+03	2430	Mouse interleukin-4 recepto
N_Geneseq_36:X22407	-	6.00	81.25	6.3e+03	2430	Murine IL-4 receptor encodi
N_Geneseq_36:T61456	-	6.00	81.17	6.4e+03	2457	C-proteinase clone pCP-1. N
N_Geneseq_36:T89399	-	6.00	81.17	6.4e+03	2458	Methods for diagnosing Frie
N_Geneseq_36:Q13311	-	6.00	81.13	6.4e+03	2469	Basic FGF receptor. Extrace
N_Geneseq_36:Q79971	-	6.00	81.09	6.4e+03	2485	Human interleukin-1 beta co
N_Geneseq_36:Q32850	+	6.00	81.08	6.4e+03	2487	BMP1. Synergistic compsn. f
N_Geneseq_36:Q31544	-	6.00	81.07	6.4e+03	2491	Murine heregulin-beta2 clone
N_Geneseq_36:Q79968	-	6.00	81.07	6.4e+03	2491	Human heregulin-1 beta co
N_Geneseq_36:V36165	-	6.00	81.04	6.5e+03	2500	Human trident transcription
N_Geneseq_36:N80631	+	6.00	81.00	6.5e+03	2514	Human Bone Morphogenic Prot
N_Geneseq_36:T91743	-	6.00	80.97	6.5e+03	2521	Lily calcium/calmodulin-dep
N_Geneseq_36:V74671	-	6.00	80.95	6.5e+03	2531	Staphylococcus aureus conti
N_Geneseq_36:Q80227	+	6.00	80.81	6.7e+03	2582	Rat NDF clone 20 DNA. New r
N_Geneseq_36:N60610	+	6.00	80.80	6.7e+03	2585	Vector sequence derived fro
N_Geneseq_36:N70529	+	6.00	80.76	6.7e+03	2599	DNA -a sequence of Bean Gol
N_Geneseq_36:X19480	-	6.00	80.72	6.7e+03	2614	Staphylococcus aureus mecB
N_Geneseq_36:X19027	-	6.00	80.71	6.7e+03	2615	Mouse serine protease BSSP-
N_Geneseq_36:Q04730	+	6.00	80.66	6.8e+03	2634	Sequence encoding glucoamyl
N_Geneseq_36:V23480	-	6.00	80.65	6.8e+03	2638	Pseudomonas OrfV secretion
N_Geneseq_36:T84164	-	6.00	80.62	6.8e+03	2650	DNA encoding an isopropylma
N_Geneseq_36:Q22831	-	6.00	80.58	6.9e+03	2662	Sequence encoding ovine int
N_Geneseq_36:Q21003	-	6.00	80.53	6.9e+03	2682	flg receptor protein gene.
N_Geneseq_36:V63471	-	6.00	80.53	6.9e+03	2682	cDNA encoding a human chron
N_Geneseq_36:V63469	-	6.00	80.52	6.9e+03	2685	cDNA encoding a human chron
N_Geneseq_36:X20667	-	6.00	80.50	6.9e+03	2692	Polynucleotide sequence fro
N_Geneseq_36:Q51114	-	6.00	80.49	6.9e+03	2698	Sequence encoding methylase
N_Geneseq_36:V53507	+	6.00	80.47	7.0e+03	2702	DNA encoding two 3-isopropy
N_Geneseq_36:V40546	-	6.00	80.44	7.0e+03	2715	Homo sapiens secreted prote
N_Geneseq_36:V64624	-	6.00	80.39	7.0e+03	2733	P. putida cis/trans isomera
N_Geneseq_36:T31051	-	6.00	80.39	7.0e+03	2733	Human fibroblast growth fac
N_Geneseq_36:T14604	-	6.00	80.39	7.0e+03	2735	Human serotonin variant 5HT
N_Geneseq_36:V48282	-	6.00	80.38	7.0e+03	2739	Human 5-HT2C serotonin rece
N_Geneseq_36:T44351	-	6.00	80.36	7.1e+03	2743	H. pylori ureA + ureB gene
N_Geneseq_36:X01067	-	6.00	80.36	7.1e+03	2743	BslI methylase gene sequenc
N_Geneseq_36:Q80231	-	6.00	80.36	7.1e+03	2745	Rat NDF clone 41 DNA. New r
N_Geneseq_36:Q14808	+	6.00	80.36	7.1e+03	2745	B.thuringiensis toxin/gp64
N_Geneseq_36:T90831	+	6.00	80.26	7.1e+03	2767	Glucoanylase P gene. Hormoc
N_Geneseq_36:Q12485	-	6.00	80.23	7.1e+03	2784	DNA encoding A and B subuni
N_Geneseq_36:T67235	-	6.00	80.16	7.2e+03	2794	Pasteurella haemolytica tra
N_Geneseq_36:T42791	+	6.00	80.15	7.2e+03	2822	Scytalidium catalase gene.
N_Geneseq_36:T51612	+	6.00			2826	Curvularia verruculosa halo
N_Geneseq_36:T38071	+					Transferrin binding protein

Left table

strand		Tm	score	num	accession	description
+	6.00	80.14	7.3e+03	2830	N_Geneseq_36:024234	Lipoxygenase gene. Prodn. of P
+	6.00	80.14	7.3e+03	2830	N_Geneseq_36:024492	Rice lipoxygenase gene. Rice l
-	6.00	80.11	7.3e+03	2842	N_Geneseq_36:026552	RING4 cDNA clone. DNA encoding
-	6.00	80.09	7.3e+03	2848	N_Geneseq_36:T86089	Kex2 protease encoding cDNA. S
-	6.00	80.07	7.3e+03	2856	N_Geneseq_36:010448	Human basic fibroblast growth
+	6.00	80.03	7.4e+03	2871	N_Geneseq_36:014807	B.thuringiensis toxin/gp64 vir
+	6.00	80.01	7.4e+03	2880	N_Geneseq_36:094642	Recombinant cold-resistant pyr
-	6.00	80.01	7.4e+03	2880	N_Geneseq_36:T12673	Corn derived pyruvate phosphat
+	6.00	79.98	7.4e+03	2891	N_Geneseq_36:057016	PKC delta. Eukaryotic cells tr
+	6.00	79.96	7.4e+03	2900	N_Geneseq_36:V65145	Ehrlichia sp. E80 genomic DNA.
-	6.00	79.96	7.4e+03	2901	N_Geneseq_36:T32961	Mouse developmental kinase 1 M
+	6.00	79.94	7.4e+03	2909	N_Geneseq_36:Q01758	cDNA sequence of rat C kinase
+	6.00	79.92	7.5e+03	2914	N_Geneseq_36:080226	Rat NDF clone 19 DNA. New recc
+	6.00	79.92	7.5e+03	2915	N_Geneseq_36:094641	Recombinant cold-resistant pyr
+	6.00	79.92	7.5e+03	2915	N_Geneseq_36:T12672	F. bidentis pyruvate phosphate
+	6.00	79.87	7.5e+03	2935	N_Geneseq_36:T05842	Calcium-independent cytosolic
+	6.00	79.87	7.5e+03	2935	N_Geneseq_36:T68827	Calcium-independent cytosolic
-	6.00	79.87	7.5e+03	2935	N_Geneseq_36:V64840	Ca-independent phospholipase A
+	6.00	79.86	7.5e+03	2935	N_Geneseq_36:Q87350	Hamster cytosolic phospholipas
-	6.00	79.86	7.5e+03	2939	N_Geneseq_36:V37293	Chinese hamster calcium indepe
-	6.00	79.86	7.5e+03	2939	N_Geneseq_36:V69280	Plasmid BGINV. New intron-medi
+	6.00	79.85	7.5e+03	2940	N_Geneseq_36:V10514	pBGINV plasmid used in the cou
+	6.00	79.85	7.6e+03	2945	N_Geneseq_36:065587	Human ITAK DNA #1. Nucleic aci
+	6.00	79.77	7.6e+03	2945	N_Geneseq_36:V69281	S. lavendulae mcr locus. Genes
-	6.00	79.75	7.6e+03	2975	N_Geneseq_36:N50494	Human chromodomain Y like (CDY
+	6.00	79.74	7.6e+03	2983	N_Geneseq_36:004023	Heat resistant aldehyde dehydr
-	6.00	79.71	7.7e+03	2990	N_Geneseq_36:095299	Gene pJH10 encoding bacterial
+	6.00	79.71	7.7e+03	3001	N_Geneseq_36:T16304	Plasmid pAIC2-26 contg. mouse
-	6.00	79.69	7.7e+03	3007	N_Geneseq_36:068842	Fas coding sequence from AIC2A
+	6.00	79.66	7.7e+03	3023	N_Geneseq_36:T67241	Plant blight-specific resistan
+	6.00	79.52	7.9e+03	3079	N_Geneseq_36:X30093	Pasteurella haemolytica transf
+	6.00	79.50	7.9e+03	3088	N_Geneseq_36:T05628	Human vascular smooth muscle c
+	6.00	79.49	7.9e+03	3091	N_Geneseq_36:N90387	ADP ribosylation factor gene r
-	6.00	79.46	7.9e+03	3106	N_Geneseq_36:V70135	Gene encoding KEX2 endoproteas
+	6.00	79.43	7.9e+03	3116	N_Geneseq_36:V70494	Pyruvate formate lyase homolog
+	6.00	79.40	7.9e+03	3133	N_Geneseq_36:090652	Eph-related tyrosine kinase CE
+	6.00	79.36	8.0e+03	3150	N_Geneseq_36:074654	Staphylococcus aureus contig S
-	6.00	79.33	8.0e+03	3161	N_Geneseq_36:080233	Rat NDF clone 42B DNA. New rec
+	6.00	79.30	8.1e+03	3175	N_Geneseq_36:074624	Staphylococcus aureus contig
+	6.00	79.29	8.1e+03	3180	N_Geneseq_36:094645	Recombinant cold-resistant pyr
-	6.00	79.29	8.1e+03	3180	N_Geneseq_36:T12676	F. brownii cold-resistant pyru
+	6.00	79.26	8.1e+03	3191	N_Geneseq_36:004092	3.2 kb KpnI-PvuII fragment of
+	6.00	79.11	8.3e+03	3258	N_Geneseq_36:V26030	Human transient receptor poten
+	6.00	79.10	8.3e+03	3264	N_Geneseq_36:V10525	Human ITAK DNA #2. Nucleic aci
+	6.00	79.05	8.3e+03	3287	N_Geneseq_36:X26587	Nucleotide sequence of human K
+	6.00	79.02	8.4e+03	3297	N_Geneseq_36:X23981	Human HG38 DNA. Human G-protei
+	6.00	79.01	8.4e+03	3303	N_Geneseq_36:V14806	B.thuringiensis toxin/gp64 vir
+	6.00	78.97	8.4e+03	3319	N_Geneseq_36:T31441	Interleukin-6 receptor coding
-	6.00	78.97	8.4e+03	3319	N_Geneseq_36:V60295	Human interleukin-6 receptor c
+	6.00	78.95	8.4e+03	3323	N_Geneseq_36:V60914	Human bFGF receptor sequence.
+	6.00	78.93	8.5e+03	3337	N_Geneseq_36:N91089	Fragment of Pseudomonas SMP1 D
-	6.00	78.92	8.5e+03	3344	N_Geneseq_36:080228	Rat NDF clone 22 DNA. New rec
+	6.00	78.87	8.5e+03	3366	N_Geneseq_36:084803	Spinocerebellar ataxia type 1
+	6.00	78.87	8.5e+03	3367	N_Geneseq_36:V63191	cDNA from clone d139_9 which e
+	6.00	78.84	8.6e+03	3368	N_Geneseq_36:T43077	Wheat acetyl-CoA carboxylase A
-	6.00	78.80	8.6e+03	3368	N_Geneseq_36:T30386	5' end flanking sequence of wh
+	6.00	78.75	8.6e+03	3382	N_Geneseq_36:Q62300	pDam-L3 plasmid fragment encod
+	6.00	78.72	8.7e+03	3401	N_Geneseq_36:Q06904	Sequence encoding site specifi
+	6.00	78.71	8.7e+03	3423	N_Geneseq_36:T00771	GST-SEP Fusion gene fragment.
+	6.00	78.70	8.7e+03	3437	N_Geneseq_36:T16273	Pyrodictium occultum 3437 bp D
+	6.00	78.67	8.7e+03	3442	N_Geneseq_36:020545	Encodes KEX2 endopeptidase wit
+	6.00	78.67	8.8e+03	3447	N_Geneseq_36:051532	M.leprae rpoB gene. Rapid dete
+	6.00	78.65	8.8e+03	3461	N_Geneseq_36:Q58005	Sequence of plasmid pTHN1 show
+	6.00	78.65	8.8e+03	3461	N_Geneseq_36:064068	Non-A, non-B hepatitis virus g
-	6.00	78.65	8.8e+03	3471	N_Geneseq_36:T93650	5'UTR/CORE/ENV/NS1/NS2/NS3' cDN
+	6.00	78.64	8.8e+03	3472	N_Geneseq_36:048496	Herpes simplex virus type 1 gl
+	6.00	78.58	8.8e+03	3472	N_Geneseq_36:Q21453	Sequence of Herpes simplex vir
-	6.00	78.58	8.8e+03	3475	N_Geneseq_36:V15221	Glycoprotein B (gB). DNA cons
-	6.00	—	8.9e+03	3501	N_Geneseq_36:V44041	Sequence encoding beta-chain C
	6.00	78.58	8.9e+03	3503		Bacillus thuringiensis 158C2b
						Mouse bFGF receptor DNA. Assay

Right table

description	accession	strand		Tm	score	num
Mitochondrial cytochrome P4	N_Geneseq_36:092829	+	6.00	78.57	8.9e+03	3510
Mouse truncated beta-3 inte	N_Geneseq_36:T61291	+	6.00	78.55	8.9e+03	3517
H6 promoted cytomegalovirus	N_Geneseq_36:T47554	-	6.00	78.54	8.9e+03	3521
Sequence encoding B.thuring	N_Geneseq_36:006830	-	6.00	78.54	8.9e+03	3522
Lepidopteran-active toxin D	N_Geneseq_36:Q10182	-	6.00	78.54	8.9e+03	3522
DNA encoding the novel toxi	N_Geneseq_36:026928	+	6.00	78.54	8.9e+03	3522
Delta endotoxin gene. Contr	N_Geneseq_36:Q47291	+	6.00	78.54	8.9e+03	3522
CryIF/436 chimeric toxin co	N_Geneseq_36:T05250	+	6.00	78.54	8.9e+03	3522
B.t. toxin 81A2 gene. Toxin	N_Geneseq_36:T16558	+	6.00	78.54	8.9e+03	3522
CryIF toxin with limited co	N_Geneseq_36:T05145	+	6.00	78.54	8.9e+03	3522
CryIF/436 chimeric toxin ge	N_Geneseq_36:T18702	+	6.00	78.54	8.9e+03	3522
Codon-reworked CryIF toxin	N_Geneseq_36:T18722	+	6.00	78.54	8.9e+03	3522
Bacillus thuringiensis PS81	N_Geneseq_36:080226	+	6.00	78.54	8.9e+03	3522
Plasmid pMYC243 Bt endotox	N_Geneseq_36:V62081	-	6.00	78.54	8.9e+03	3522
Plasmid pMYC2254 cryIF/436	N_Geneseq_36:V62083	+	6.00	78.54	8.9e+03	3522
Enterococcus faecalis genom	N_Geneseq_36:X13437	+	6.00	78.54	8.9e+03	3522
N-sam cDNA. Gene prod of N-	N_Geneseq_36:Q27658	-	6.00	78.54	8.9e+03	3525
Gene encoding bacterial tox	N_Geneseq_36:004284	-	6.00	78.52	8.9e+03	3531
Delta endotoxin gene. Contr	N_Geneseq_36:Q47290	+	6.00	78.52	8.9e+03	3531
CryIE(d) gene. New DNA enco	N_Geneseq_36:T09329	+	6.00	78.52	8.9e+03	3531
Bacillus thuringiensis var.	N_Geneseq_36:V00293	-	6.00	78.52	8.9e+03	3531
Bacillus thuringiensis EG11	N_Geneseq_36:V37293	+	6.00	78.52	8.9e+03	3531
Bacillus thuringiensis EG11	N_Geneseq_36:V31160	+	6.00	78.52	8.9e+03	3531
Bacillus thuringiensis EG11	N_Geneseq_36:V31161	+	6.00	78.52	8.9e+03	3531
Human inhibitor of apoptosi	N_Geneseq_36:T72711	+	6.00	78.52	8.9e+03	3532
Encodes modified B.t.k. HD-	N_Geneseq_36:Q05794	-	6.00	78.52	8.9e+03	3534
Encodes full-length modifie	N_Geneseq_36:Q05795	+	6.00	78.52	8.9e+03	3534
Encodes full-length endocti	N_Geneseq_36:006561	+	6.00	78.52	8.9e+03	3534
Bacillus thuringiensis EG11	N_Geneseq_36:V31162	+	6.00	78.52	8.9e+03	3534
Bacillus thuringiensis EG11	N_Geneseq_36:V31163	+	6.00	78.52	8.9e+03	3534
Bacillus thuringiensis EG11	N_Geneseq_36:005797	-	6.00	78.51	8.9e+03	3534
Bacillus thuringiensis CryI	N_Geneseq_36:Q56806	-	6.00	78.51	8.9e+03	3537
FIPV spike (S) gene C-termi	N_Geneseq_36:T69981	+	6.00	78.51	8.9e+03	3537
Homo sapiens pcdh-4 cDNA.	N_Geneseq_36:V27594	+	6.00	78.50	8.9e+03	3540
Eph-related PTK Cek10 cDNA.	N_Geneseq_36:Q90656	-	6.00	78.49	9.0e+03	3546
C-proteinase clone pCP-2. N	N_Geneseq_36:T61457	+	6.00	78.49	9.0e+03	3546
Gene encoding CryIE/CryIC B	N_Geneseq_36:Q80648	+	6.00	78.47	9.0e+03	3558
Human TYK2 kinase coding se	N_Geneseq_36:Q85414	-	6.00	78.46	9.0e+03	3561
Human c-fos oncogene. Antic	N_Geneseq_36:V20466	-	6.00	78.45	9.0e+03	3565
Encodes B.t.entomocidus ins	N_Geneseq_36:005797	+	6.00	78.45	9.0e+03	3567
Lepidopteran-active toxin C	N_Geneseq_36:Q10181	-	6.00	78.45	9.0e+03	3567
Delta endotoxin gene. Contr	N_Geneseq_36:Q47294	+	6.00	78.45	9.0e+03	3567
CryIC gene. New Bacillus th	N_Geneseq_36:Q80646	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27845	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27846	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27847	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27848	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27844	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27863	+	6.00	78.45	9.0e+03	3567
DNA encoding a mutated Cry1	N_Geneseq_36:V27864	+	6.00	78.45	9.0e+03	3567
Gene encoding CryIA/CryIC B	N_Geneseq_36:Q80648	+	6.00	78.42	9.0e+03	3579
Bacillus thuringiensis EG11	N_Geneseq_36:V31164	+	6.00	78.40	9.0e+03	3579
Eph-related PTK Cek10+ cDNA	N_Geneseq_36:Q90658	-	6.00	78.30	9.1e+03	3591
Cystathionine gamma synthas	N_Geneseq_36:T63514	+	6.00	77.99	9.2e+03	3639
Human prostate protein HPA3	N_Geneseq_36:X28141	+	6.00	77.97	9.2e+03	3642
Staphylococcus aureus cont1	N_Geneseq_36:T84934	+	6.00	77.91	9.3e+03	3660
Tobacco necrosis virus stra	N_Geneseq_36:V74483	+	6.00	77.90	9.3e+03	3684
Plasmid pBSGFP expressing g	N_Geneseq_36:V13115	-	6.00	77.86	9.3e+03	3699
Sequence of a novel hybrid	N_Geneseq_36:V14340	+	6.00	77.85	9.4e+03	3718
Nop-B, non-C, non hepatit	N_Geneseq_36:N90339	-	6.00	77.84	9.4e+03	3722
Human HIAP-2 coding sequenc	N_Geneseq_36:X18977	+	6.00	77.81	9.4e+03	3732
Mycobacterium urease gene c	N_Geneseq_36:V55040	+	6.00	78.12	9.4e+03	3789
CCoAOMT promoter. PtCCoAOMT	N_Geneseq_36:063514	-	6.00	78.01	9.5e+03	3800
P40 genomic DNA. Nucleic ac	N_Geneseq_36:X28141	+	6.00	77.99	9.6e+03	3809
ILTV genome fragment. Live	N_Geneseq_36:Q12760	+	6.00	77.97	9.6e+03	3841
P. suis leukotoxin genes ps	N_Geneseq_36:V07010	+	6.00	77.91	9.6e+03	3848
Homo sapiens pcdh-4 gene.	N_Geneseq_36:V27593	+	6.00	77.90	9.7e+03	3868
Phospholipase C-gamma-1 cDN	N_Geneseq_36:T12292	+	6.00	77.86	9.7e+03	3870
Leukotoxin genes hlyB and h	N_Geneseq_36:T12293	+	6.00	77.85	9.7e+03	3879
Phospholipase C-gamma-1 cDN	N_Geneseq_36:046410	+	6.00	77.84	9.7e+03	3879
Sequence encoding incectici	N_Geneseq_36:T12293	-	6.00	77.81	9.8e+03	3893
Human mcl-1 gene. New myelo	N_Geneseq_36:Q10229	+	6.00	77.75	9.8e+03	3923
	N_Geneseq_36:Q80523	-	6.00	77.71	9.9e+03	3946

```
N_Geneseq_36:V19972   +   6.00   77.66   1.0e+04   3976   ! Homo sapiens cdo tumour suppre
N_Geneseq_36:Q15153   +   6.00   77.63   1.0e+04   3993   ! Swine herpes virus-1 major cap
N_Geneseq_36:V24310   +   6.00   77.62   1.0e+04   3997   ! Homo sapiens Class II tumour s
N_Geneseq_36:V26082   +   6.00   77.62   1.0e+04   3997   ! Tomato pest resistance M1 gene
N_Geneseq_36:X29857   -   6.00   77.61   1.0e+04   4001   ! Thermotoga neapolitana type F
N_Geneseq_36:T09328   -   6.00   77.61   1.0e+04   4003   ! CryIE(c) gene. New DNA encodin
N_Geneseq_36:X13197   -   6.00   77.59   1.0e+04   4013   ! Enterococcus faecalis genome c
N_Geneseq_36:Q14648   -   6.00   77.58   1.0e+04   4020   ! cryIF gene. Purifying and isol
N_Geneseq_36:Q52674   -   6.00   77.56   1.0e+04   4029   ! F-spondin (FP5-9) coding seque
N_Geneseq_36:X30095   -   6.00   77.56   1.0e+04   4029   ! Rat vascular smooth muscle cel
N_Geneseq_36:N90382   -   6.00   77.54   1.0e+04   4038   ! Genes encoding human alpha-2-p
N_Geneseq_36:Q15154   +   6.00   77.53   1.0e+04   4044   ! Plasmid pHAS2-MCP containing s
```

seq_name: N_Geneseq_36:V74878

```
seq_documentation_block:
ID   V74878 standard; DNA; 676 BP.
AC   V74878;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #567.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key       Location/Qualifiers
FT   misc_feature    421..480
FT           /*tag= a
FT           /note= "these bases represent a line of missing text in
FT                   the sequence listing in the specification. They
FT                   are included to maintain the nucleotide numbering
FT                   given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1505-1506; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence  676 BP;  163 A;  136 C;  77 G;  236 T;

alignment_scores:
       Quality:   10.00              Length:     10
       Ratio:     1.000              Gaps:        0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V74878/rev  ..
```

```
Align seg 1/1  to reverse of: V74878  from: 1  to: 676

176 GluGlyArgAsnArgGlnValArgArgMet 185
    ||||||||||||||||||||||||||||||
174 GAAGGTCGTAATCGCCAAGTCAGAAGAATG 145
```

seq_name: N_Geneseq_36:X20509

```
seq_documentation_block:
ID   X20509 standard; DNA; 12838 BP.
AC   X20509;
DT   05-MAY-1999  (first entry)
DE   Polynucleotide sequence from the genome of Treponema pallidum.
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
OS   Treponema pallidum.
PN   WO9859034-A2.
PD   30-DEC-1998.
PF   23-JUN-1998; U13041.
PR   24-JUN-1997; US-050667.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Fraser CM;
DR   WPI; 99-081273/07.
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
PS   Claim 1; Page 230-237; 1150pp; English.
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
SQ   Sequence  12838 BP;  3171 A;  4046 C;  2834 G;  2770 T;

alignment_scores:
       Quality:    8.00              Length:      8
       Ratio:     1.000              Gaps:        0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X20509  ..

Align seg 1/1  to: X20509  from: 1  to: 12838

  55 GlnPheThrAspGluAlaGlyAla 62
     ||||||||||||||||||||||||
7113 CAGTTCACTGACGAAGCTGGGCGC 7136
```

seq_name: N_Geneseq_36:Q69432

```
seq_documentation_block:
ID   Q69432 standard; DNA; 46 BP.
AC   Q69432;
DT   27-FEB-1995  (first entry)
DE   Human heat shock protein (hsp70B) gene, target region.
KW   DNA protein-binding assay; test sequence; screening sequence;
KW   promoter; target; TATA box; Herpes simplex virus; HSV;
KW   origin of replication; UL9; transcription factor; TFIID; ds.
OS   Synthetic.
PN   WO9414980-A.
PD   07-JUL-1994.
PF   20-DEC-1993; U12388.
PR   23-DEC-1992; US-996783.
PR   17-SEP-1993; US-123936.
PA   (GENE-) GENELABS TECHNOLOGIES INC.
PI   Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR   WPI; 94-234711/28.
PT   Sequence-directed DNA-binding molecules - useful in
PT   pharmaceuticals and as molecular reagents
PS   Claim 28; Page 303; 587pp; English.
CC   A DNA protein-binding assay is provided, useful for screening
```

CC libraries of synthetic or biological cpds. for their ability
CC to bind DNA test sequences. The assay is versatile in that any
CC number of test sequences can be tested by placing the test sequence
CC adjacent to a defined protein-binding screening sequence. Binding
CC of mols. to these test sequences changes the binding characteristics
CC of the protein mol. to its cognate binding sequence. When such a mol.
CC binds the test sequence, the equilibrium of the DNA:protein complexes
CC is disturbed, generating changes in the concentration of free DNA probe.
CC One application of this method is to eukaryotic general transcription
CC factors (e.g. TFIID), where the target region is typically selected
CC from DNA sequences adjacent to the binding site for the eukaryotic
CC transcription factor. Numerous exemplary test sequences are given:
CC the sequences in Q69251-731 and Q69850 correspond to promoter targets
CC (typically, TATA box-contg. sites) for human genes and the sequences in
CC Q69732-849 correspond to promoter targets for viral genes. The test
CC sequences may also be randomly generated. DNA:protein interaction may
CC be used for screening purposes, e.g. the Herpes Simplex Virus (HSV)
CC origin of replication and Ul9 (see Q69851-52, Q69865 and Q69891).
SQ Sequence 46 BP;   4 A;   18 C;   20 G;   4 T;

alignment_scores:
        Quality: 7.00   Length: 7
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x Q69432

Align seg 1/1 to: Q69432 from: 1   to: 46

  98 AlaArgLeuThrGlnProGly 104
     |||||||||||||||||||||
   3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: N_Geneseq_36:T63894

seq_documentation_block:
ID   T63894 standard: DNA; 46 BP.
AC   T63894;
DT   14-MAR-1997 (first entry)
DE   Human hsp70B gene TFIID binding site.
KW   Duplex DNA; target region; binding characteristic; DNA binding protein;
KW   TFIID; transcription factor; binding site; inhibition; enhance; hCG;
KW   cancer; inherited genetic disorder; alpha-D-galactosidase A; ds.
OS   Homo sapiens.
PN   US5578444-A.
PD   26-NOV-1996.
PF   27-JUN-1991; 723618.
PR   27-JUN-1991; US-723618.
PR   23-DEC-1992; US-996783.
PR   17-SEP-1993; US-123936.
PR   20-DEC-1993; US-171389.
PA   (GENE-) GENELABS TECHNOLOGIES INC.
PI   Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR   WPI; 97-020402/02.
PT   Altering binding characteristics of DNA binding proteins to duplex
PT   DNA - by attaching specific small cpd. to target region close to the
PT   protein's binding site, useful in treatment of viral disease, cancer
PT   etc
PS   Claim 6; Column 191-192; 264pp; English.
CC   The sequences given in T63713-4312 represent duplex DNA's which act
CC   as target regions in the method of the invention. The method for
CC   altering the binding characteristics of a DNA-binding protein to duplex
CC   DNA comprises contacting the duplex DNA with a small molecule which
CC   binds sequence-specifically to a target region, where, when the small
CC   molecule is bound to the target region, it is adjacent to, but not
CC   overlapping by more than 4 bp, a binding site for a DNA-binding protein.
CC   The small molecule is added at a concentration effective to alter the
CC   binding of the DNA binding protein, pref. TFIID, to its binding site on
CC   the duplex DNA. The binding of the small molecule may inhibit or enhance
CC   the binding of the DNA-binding protein to its binding site. The
CC   compounds isolated using this method are potentially useful as ...

CC therapeutic agents for treatment of any disease which involves a
CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
CC The method is suitable for screening large biological or chemical
CC libraries and allows determination of sequence-specific and relative
CC affinities of known DNA-binding agents for different DNA sequences.
CC The design of these duplex DNA's allows a single DNA:protein interaction
CC to be used for screening sequence-specific, or preferential, DNA binding
CC proteins that recognise almost any possible sequence (see also T49539-
CC 74).
SQ Sequence 46 BP;   4 A;   18 C;   20 G;   4 T;

alignment_scores:
        Quality: 7.00   Length: 7
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x T63894

Align seg 1/1 to: T63894 from: 1   to: 46

  98 AlaArgLeuThrGlnProGly 104
     |||||||||||||||||||||
   3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: N_Geneseq_36:X17182

seq_documentation_block:
ID   X17182 standard: DNA; 46 BP.
AC   X17182;
DT   06-MAY-1999 (first entry)
DE   Test sequence from human hsp70B gene.
KW   Test sequence; DNA-binding molecule; screening sequence; human;
KW   nucleic acid amplification; target; viral; ds.
OS   Homo sapiens.
PN   US5869241-A.
PD   09-FEB-1999.
PF   07-JUN-1995; 475228.
PR   20-DEC-1993; US-171389.
PR   27-JUN-1991; US-723618.
PR   23-DEC-1992; US-996783.
PR   17-SEP-1993; US-123936.
PR   07-JUN-1995; US-475228.
PA   (GENE-) GENELABS TECHNOLOGIES INC.
PI   Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
DR   WPI; 99-152755/13.
PT   Determination of DNA sequence preference of a DNA-binding molecule -
PT   based on inhibition of binding of protein to oligonucleotide
PT   sequence attached to test sequence
PS   Claim 3; Columns 193-194; 270pp; English.
CC   Sequences X17001 to X17600 represent specifically claimed target test
CC   sequences that are used in the method of the invention of determining the
CC   DNA sequence preference of a DNA-binding molecule. The method comprises:
CC   (i) adding a test molecule and a DNA-binding protein to a mixture of
CC   duplex DNA test oligonucleotides, each of the test oligonucleotides
CC   having a test sequence adjacent to a screening sequence, where the
CC   screening sequence binds to the DNA-binding protein with a binding
CC   affinity that is independent of the DNA sequence of the test sequence,
CC   and where the mixture of duplex DNA test oligonucleotide includes
CC   several test sequences; (ii) incubating the test molecule, the mixture of
CC   duplex DNA test oligonucleotides and the DNA-binding protein for a time
CC   sufficient to permit binding of the test molecule to test sequences in
CC   the duplex DNA; (iii) separating unbound test oligonucleotides from test
CC   oligonucleotides bound to binding protein; (iv) amplifying the unbound
CC   test oligonucleotides; (v) repeating steps (ii) to (iv); (vi) isolating
CC   the amplified test oligonucleotides; and (vii) sequencing the isolated
CC   test oligonucleotide. Test sequences X17001-X17481 and X17600 correspond
CC   to promoter targets for human genes and test sequences X17482-X17599
CC   correspond to promoter targets for viral genes.
SQ   Sequence 46 BP;   4 A;   18 C;   20 G;   4 T;

```
alignment_scores:
    Quality:    7.00          Length:        7
    Ratio:      1.000         Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X17182    ..

Align seg 1/1  to: X17182  from: 1  to: 46

      98 AlaArgLeuThrGlnProGly 104
         ||||||||||||||||||||||
       3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: N_Geneseq_36:V21143

seq_documentation_block:
ID  V21143 standard; DNA; 286 BP.
AC  V21143;
DT  30-JUL-1998 (first entry)
DE  Human hsp-70B promoter.
KW  Heat shock promoter; therapeutic; tumour; ss.
OS  Homo sapiens.
PN  WO9806864-A2.
PD  19-FEB-1998.
PF  14-AUG-1997; U15270.
PR  15-AUG-1996; US-024213.
PA  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI  Moonen C;
DR  WPI; 98-168793/15.
PT  Spatial and temporal control of construct gene by heat shock
PT  promoter - for production of therapeutic protein in cell mass,
PT  especially for treating tumours
PS  Claim 3; Fig 3c; 58pp; English.
CC  The Human Hsp-70B promoter and its analogue from Drosophila (V21142) can
CC  be activated using localised heat.  The hsp 70B heat shock gene promoter
CC  was used to control expression of a gene construct.  This can be used for
CC  the expression and production of therapeutic proteins in a preselected
CC  region of a cell mass.  This is especially useful for the targeted
CC  expression of therapeutic proteins in a tumour such as a prostate tumour.
SQ  Sequence   286 BP;   44 A;   91 C;  109 G;   42 T;

alignment_scores:
    Quality:    7.00          Length:        7
    Ratio:      1.000         Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V21143    ..

Align seg 1/1  to: V21143  from: 1  to: 286

      98 AlaArgLeuThrGlnProGly 104
         ||||||||||||||||||||||
     163 GCCCGGCTGACTCAGCCCGGG 183

seq_name: N_Geneseq_36:V49571

seq_documentation_block:
ID  V49571 standard; cDNA to mRNA; 447 BP.
AC  V49571;
DT  21-OCT-1998 (first entry)
DE  Human lymphoma cell line U937 clone HP10085 cDNA #2.
KW  Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW  differentiation; immune system; stimulator; suppressor; regulator;
KW  heamatopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW  haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
OS  Homo sapiens.
PN  WO9821328-A2.
PD  22-MAY-1998.
PF  07-NOV-1997; J04056.
PR  13-NOV-1996; JP-301429.
PA  (PROT-) PROTEGENE INC.
PA  (SAGA ) SAGAMI CHEM RES CENTRE.
PI  Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR  WPI; 98-297932/26.
DR  P-PSDB; W64544.
PT  Human protein having transmembrane domain - useful for, e.g.
PT  research and nutrition
PS  Claim 3; Page 124; 205pp; English.
CC  V49550-V49599 are cDNA sequences which encode human proteins containing a
CC  transmembrane domain. These proteins can be used for, e.g. research and
CC  nutrition, and may have cytokine and cell proliferation/differentiation,
CC  immune stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC  activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC  receptor/ligand, anti-inflammatory or tumour inhibition activity.
SQ  Sequence   447 BP;  161 A;   70 C;   93 G;  123 T;

alignment_scores:
    Quality:    7.00          Length:        7
    Ratio:      1.000         Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V49571    ..

Align seg 1/1  to: V49571  from: 1  to: 447

     216 AspValThrProLysGluAsn 222
         ||||||||||||||||||||||
     404 GATGTTACACCGAAAGAAAAT 424

seq_name: N_Geneseq_36:V24918

seq_documentation_block:
ID  V24918 standard; DNA; 579 BP.
AC  V24918;
DT  30-JUN-1998 (first entry)
DE  H. pylori cytoplasmic protein ORF 04ge10816_22086531_f2.10.
KW  Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW  identification; binding compound; bacteria; life cycle; activator;
KW  inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW  bacterium; ds.
OS  Helicobacter pylori.
FH  Key             Location/Qualifiers
FT  CDS             1..579
FT                  /*tag=  a
PN  WO9737044-A1.
PD  09-OCT-1997.
PF  27-MAR-1997; U05223.
PR  06-DEC-1996; US-761318.
PR  29-MAR-1996; US-625811.
PR  02-APR-1996; US-758731.
PR  25-OCT-1996; US-758711.
PR  28-OCT-1996; US-738859.
PA  (ASTR ) ASTRA AB.
PI  Alm RA, Smith D;
DR  WPI; 97-503122/46.
DR  P-PSDB; W5509.
PT  Helicobacter pylori nucleic acid sequences and encoded
PT  polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT  infection and for diagnosis of H. pylori infection
PS  Claims 5,6,42; Page 327; 1145pp; English.
CC  This sequence encodes a H. pylori cytoplasmic protein.
CC  The protein may be used in a vaccine to prevent or treat H. pylori
CC  infection or to identify H. pylori polypeptide binding compounds,
CC  useful as potential H. pylori life cycle activators or inhibitors.
CC  The DNA and probes derived from it may be used for the
CC  identification of H. pylori in a sample, and the diagnosis of
CC  H. pylori infection. Nucleic acid sequences complementary to the
CC  DNA act as antisense sequences, and can be used to prevent the
CC  translation of H. pylori mRNA. Antibodies against the protein can
CC  be used in immunoassays to evaluate the abundance and distribution
CC  of H. pylori-specific antigens. The genomic sequence of H. pylori
```

CC (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were
CC analysed for ORF of at least 180 nucleotides, and the predicted
CC coding regions defined by computer evaluation. To identify likely
CC H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having
CC identified and determined the sequences of interest, particular
CC regions can be isolated from H. pylori by PCR amplification for
CC recombinant polypeptide production, e.g. in E. coli hosts.
SQ Sequence 579 BP; 153 A; 112 C; 140 G; 174 T;

alignment_scores:
        Quality: 7.00   Length: 7
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V24918

Align seg 1/1 to: V24918 from: 1 to: 579

128 ArgAsnGlyValThrLeuAsn 134
    |||||||||||||||||||||
10 AGGAATGGCGTTACTTTAAAT 30

seq_name: N_Geneseq_36:T01289

seq_documentation_block:
ID T01289 standard; cDNA; 613 BP.
AC T01289;
DT 08-JUL-1996 (first entry)
DE Rhipicephalus appendiculatus protective antigen IGBP-MC cDNA.
KW Male tick; protective antigen; metazoan parasite; vaccine; helminth;
KW arthropod; fly; mite; flea; immunoglobulin binding protein; IGBP;
OS Rhipicephalus appendiculatus
FH Key   Location/Qualifiers
FT cds   1..519
FT   /*tag= a
FT misc_feature 61..120
FT   /*tag= b
FT   /note= "identical in N-terminal/translated sequence"
PN WO9527056-A1.
PD 12-OCT-1995.
PF 04-APR-1995; G00772.
PR 05-APR-1994; GB-006708.
PA (MLCW) MALLINCKRODT VETERINARY INC.
PI Nuttall PA, Wang H;
DR WPI; 95-35863/46.
DR P-PSDB; R82811.
PT New metazoan parasite antigens - which bind to host immunoglobulin,
PT used to prepare vaccines for protection against metazoan parasites
PS Claim 23; Fig.27C; 93pp; English.
CC This cDNA encodes IGBP-MC (an immunoglobulin binding protein) from
CC the male tick. This sequence may be expressed recombinantly for the
CC production of the protein which is capable of binding host Ig as
CC a vaccine target, thereby depriving the parasite of an important
CC means of defence against the immune response of the host. The
CC encoded antigenic protein can be used in vaccines for stimulating an
CC immune response against metazoan paraistes in humans or animals.
CC They can be used for protection against helminths and arthropod
CC parasites e.g. flies, ticks, mites, fleas or bugs.
SQ Sequence 613 BP; 169 A; 138 C; 166 G; 140 T;

alignment_scores:
        Quality: 7.00   Length: 7
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:

US-09-252-691-7056 x T01289/rev

Align seg 1/1 to reverse of: T01289   from: 1 to: 613

30 GlnAlaThrArgArgThrPro 36
   |||||||||||||||||||||
34 CAAGCAACACGGCGGACACCG 14

seq_name: N_Geneseq_36:V24773

seq_documentation_block:
ID V24773 standard; DNA; 624 BP.
AC V24773;
DT 17-JUN-1998 (first entry)
DE H. pylori ORF 04gell210orf1.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
OS Helicobacter pylori.
FH Key   Location/Qualifiers
FT CDS   1..624
FT   /*tag= a
FT   /note= "no stop codon given"
PN WO9737044-A1.
PD 09-OCT-1997.
PF 27-MAR-1997; U05223.
PR 06-DEC-1996; US-761318.
PR 29-MAR-1996; US-625811.
PR 02-APR-1996; US-758731.
PR 25-OCT-1996; US-736905.
PR 28-OCT-1996; US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI; 97-503122/46.
DR P-PSDB; W55364.
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 5,6; Page 236; 1145pp; English.
CC This sequence encodes a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
SQ Sequence 624 BP; 166 A; 129 C; 148 G; 181 T;

alignment_scores:
        Quality: 7.00   Length: 7
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V24773

Align seg 1/1 to: V24773 from: 1 to: 624

128 ArgAsnGlyValThrLeuAsn 134

```
                    |||||||||||||||||||||||
                 58 AGGAATGGCGTTACTTTAAAT 78
```

seq_name: N_Geneseq_36:X21163

seq_documentation_block:
ID X21163 standard; DNA; 641 BP.
AC X21163;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 1101; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 641 BP; 151 A; 217 C; 180 G; 92 T;
```

```
alignment_scores:
        Quality:   7.00              Length:       7
        Ratio:     1.000             Gaps:         0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X21163

Align seg 1/1  to: X21163  from: 1  to: 641

   35  ThrProGluProGlnProThr 41
       |||||||||||||||||||||
   77  ACGCCCGAGCCCCAGCCGACA 97
```

seq_name: N_Geneseq_36:X15049

seq_documentation_block:
ID X15049 standard; DNA; 654 BP.
AC X15049;
DT 15-APR-1999 (first entry)
DE DNA encoding a GST-MIG fusion protein.
KW Monokine induced by gamma-interferon; MIG; CXC chemokine; metastasis;
KW angiogenesis inhibitor; angiogenesis inducer; tumour growth inhibition;
KW haemangiomas; rheumatoid arthritis; atherosclerosis; meningioma;
KW idiopathic pulmonary fibrosis; benign prostatic hypertrophy; psoriasis;
KW vascular restenosis; arteriovenous malformation; neovascular glaucoma;
KW angiofibroma; haemophilic joint; hypertrophic scar; Osler-Weber syndrome;
KW pyogenic granuloma retrolental fibroplasia; scleroderma; trachoma;
KW vascular adhesion; synovitis; dermatitis; endometriosis; pterygium;
KW diabetic retinopathy; neovascularisation; chronic bronchitis;
KW adult respiratory distress syndrome; ARDS; pseudogout;
KW cystic fibrosis; fusion protein; ss.
OS Synthetic.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   1..654
FT   /tag= a
FT   /note= "sequence contains in-frame insertions"
PN US5871723-A.
PD 16-FEB-1999.
PF 06-JUN-1995; 468819.
PR 06-JUN-1995; US-468819.
PA (UNMI) UNIV MICHIGAN.
DR WPI; 99-166569/14.
DR P-PSDB: W96736.
PT Use of chemokines with a conserved Cys Xaa Cys (CXC) sequence -
PT which do not contain amino acid sequence ELR, for inhibiting
PT angiogenesis in tumours, rheumatoid arthritis, restenosis or
PT glaucoma
PS Example 13; Columns 127-128; 145pp; English.
CC The present sequence encodes a glutathione-S-transferase (GST)-monokine
CC induced by gamma-interferon (MIG) fusion protein. MIG is a CXC chemokine
CC that is an inhibitor of angiogenesis. The specification describes methods
CC for inhibiting angiogenesis or for inducing angiostasis, using chemokines
CC (with a conserved Cys Xaa Cys (CXC) sequence at the N-terminal) other
CC than platelet factor-4, and which do not contain the amino acid sequence
CC ELR. The methods are useful for inhibiting tumour growth and metastasis
CC and for treating diseases such as haemangiomas, rheumatoid arthritis,
CC atherosclerosis and idiopathic pulmonary fibrosis (IPF), benign prostatic
CC hypertrophy (BPH), vascular restenosis, arteriovenous malformation
CC (AVM), meningioma, neovascular glaucoma, psoriasis, angiofibroma,
CC haemophilic joints, hypertrophic scars, Osler-Weber syndrome, pyogenic
CC granuloma retrolental fibroplasia, scleroderma, trachoma, vascular
CC adhesions, synovitis, dermatitis, endometriosis, pterygium, diabetic
CC retinopathy neovascularisation associated with corneal injury or grafts,
CC adult respiratory distress syndrome (ARDS), chronic bronchitis,
CC pseudogout and cystic fibrosis.
SQ Sequence 654 BP; 192 A; 135 C; 135 G; 192 T;
```

```
alignment_scores:
        Quality:   7.00              Length:       7
        Ratio:     1.000             Gaps:         0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X15049/rev

Align seg 1/1  to reverse of: X15049  from: 1  to: 654

  157  ProProIleArgGluArgLys 163
       |||||||||||||||||||||
  196  CCACCTATCAGGGAACGGAAG 176
```

seq_name: N_Geneseq_36:X15057

seq_documentation_block:
ID X15057 standard; DNA; 654 BP.
AC X15057;
DT 15-APR-1999 (first entry)
DE DNA encoding GST-MIG fusion protein.
KW monokine induced by gamma-interferon; MIG; CXC chemokine; metastasis;
KW angiogenesis inhibitor; angiogenesis inducer; tumour growth inhibition;
KW haemangiomas; rheumatoid arthritis; atherosclerosis; meningioma;
KW idiopathic pulmonary fibrosis; benign prostatic hypertrophy; psoriasis;
KW vascular restenosis; arteriovenous malformation; neovascular glaucoma;
KW angiofibroma; haemophilic joint; hypertrophic scar; Osler-Weber syndrome;
KW pyogenic granuloma retrolental fibroplasia; scleroderma; trachoma;
KW vascular adhesion; synovitis; dermatitis; endometriosis; pterygium;
KW diabetic retinopathy; neovascularisation; chronic bronchitis;
KW adult respiratory distress syndrome; ARDS; pseudogout;
KW cystic fibrosis; fusion protein; ss.
OS Synthetic.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   1..627
FT   /tag= a
FT   /note= "sequence contains numerous in-frame
FT   insertions"
PN US5871723-A.
PD 16-FEB-1999.
PF 06-JUN-1995; 468819.
```

```
PR  06-JUN-1995; US-468819.
PA  (UNMI ) UNIV MICHIGAN.
PI  Kunkel SL, Polverini PJ, Strieter RM;
DR  WPI; 99-166569/14.
DR  P-PSDB; W96740.
PT  Use of chemokines with a conserved Cys Xaa Cys (CXC) sequence -
PT  which do not contain amino acid sequence ELR, for inhibiting
PT  angiogenesis in tumours, rheumatoid arthritis, restenosis or
PT  glaucoma
PS  Example 16; Columns 129-130; 145pp; English.
CC  The present sequence encodes a glutathione-S-transferase (GST)-monokine
CC  induced by gamma-interferon (MIG) fusion protein. MIG is a CXC chemokine
CC  that is an inhibitor of angiogenesis. The specification describes methods
CC  for inhibiting angiogenesis or for inducing angiostasis, using chemokines
CC  (with a conserved Cys Xaa Cys (CXC) sequence at the N-terminal) other
CC  than platelet factor-4, and which do not contain the amino acid sequence
CC  ELR. The methods are useful for inhibiting tumour growth and metastasis
CC  and for treating diseases such as haemangiomas, rheumatoid arthritis,
CC  atherosclerosis and idiopathic pulmonary fibrosis (IPF), benign prostatic
CC  hypertrophy (BPH), vascular restenosis, arteriovenous malformations
CC  (AVM), meningioma, neovascular glaucoma, psoriasis, angiofibroma,
CC  haemophilic joints, hypertrophic scars, Osler-Weber syndrome, pyogenic
CC  granuloma retrolental fibroplasia, scleroderma, trachoma, vascular
CC  adhesions, synovitis, dermatitis, endometriosis, pterygium, diabetic
CC  retinopathy neovascularisation associated with corneal injury or grafts,
CC  adult respiratory distress syndrome (ARDS), chronic bronchitis,
CC  pseudogout and cystic fibrosis.
SQ  Sequence  654 BP;  192 A;  135 C;  135 G;  192 T;

alignment_scores:
         Quality:    7.00           Length:    7
         Ratio:      1.000          Gaps:      0
    Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X15057/rev   ..

Align seg 1/1  to reverse of: X15057  from: 1  to: 654

157 ProProIleArgGluArgLys 163
    |||||||||||||||||||||
196 CCACCTATCAGGGAACGGAAG 176

seq_name: N_Geneseq_36:V83108

seq_documentation_block:
ID  V83108 standard; DNA; 655 BP.
AC  V83108;
DT  02-MAR-1999  (first entry)
DE  Human C-type lectin-1 (CTL-1) coding sequence.
KW  lectin; carbohydrate; binding; agglutination; selectins; receptors;
KW  calcium dependent binding; treatment; cancer; detection;
KW  identification; CTL-1; ovary; paraganglionic; penis; brain; thyroid;
KW  heart; ss.
OS  Homo sapiens.
FH  Key          Location/Qualifiers
FT  CDS          163..612
FT               /tag= a
FT               /product= C-type_lectin_(CTL)-1
PN  WO9849306-A1.
PD  05-NOV-1998.
PF  29-APR-1998; U08791.
PR  29-APR-1997; US-846523.
PA  (INCY-) INCYTE PHARM INC.
PI  Bandman O, Shah P;
DR  WPI; 99-024060/02.
DR  P-PSDB; W85592.
PT  New human C-type lectin and related nucleic acid, vectors,
PT  transformed cells - antibodies, agonists and antagonists, for
PT  diagnosis, prevention and treatment of cancers
PS  Claim 5; Page 47; 64pp; English.
CC  Recombinant cells containing expression vectors comprising this CTL-1
CC  coding sequence can be used to produce recombinant CTL-1 which is
CC  useful for raising Ab and to screen for specific binding agents.
CC  Binding agents which are antagonists of CTL-1 can be used to
CC  treat or prevent cancer, e.g. of ovary, paraganglion, penis, brain,
CC  thyroid and heart which are associated with overexpression of
CC  CTL-1. Fragments of this sequence are used to diagnose these
CC  conditions, as probes or primers in usual hybridisation and/or
CC  amplification assays, or for gene mapping, while complements of
CC  this sequence, antisense or ribozyme sequences are used to treat or
CC  prevent the aforementioned cancers also. Ab are used directly as
CC  antagonists or for delivery of therapeutic agents to cells that
CC  express CTL-1. CTL-1 includes a carbohydrate-recognition domain
CC  (aa35-149) and also a transmembrane domain, but no secretory signal.
CC  CTL-1 shows homology with human CD69 (W85593), chicken 17.5.3 (W85594),
CC  and mouse CD69 (W85595) proteins.
SQ  Sequence  655 BP;  247 A;  96 C;  132 G;  180 T;

alignment_scores:
         Quality:    7.00           Length:    7
         Ratio:      1.000          Gaps:      0
    Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x V83108   ..

Align seg 1/1  to: V83108  from: 1  to: 655

216 AspValThrProLysGluAsn 222
    |||||||||||||||||||||
566 GATGTTACCACCGAAAGAAAT 586

seq_name: N_Geneseq_36:Q86546

seq_documentation_block:
ID  Q86546 standard; DNA; 667 BP.
AC  Q86546;
DT  23-NOV-1995  (first entry)
DE  Wilson disease exon 5/intron borders.
KW  Copper transport; ATPase; Wilson disease; chromosomal markers; probe;
KW  yeast artificial chromosome library; hybridisation; Menkes disease;
KW  copper binding site; phosphate domain; transduction; transmembrane;
KW  phosphorylation; ATP binding domain; autosomal recessive disorder;
KW  toxicity; liver; brain; gene therapy; heavy metal; exon; intron; ss.
OS  Homo sapiens.
FH  Key          Location/Qualifiers
FT  intron       1..312
FT               /tag= a
FT               /anticodon=
FT  exon         313..389
FT               /tag= b
FT               /note= "exon 5"
FT  intron       390..667
FT               /tag= c
PN  WO9508641-A.
PD  30-MAR-1995.
PF  21-SEP-1994; CA0519.
PR  21-SEP-1993; CA-106602.
PR  21-OCT-1993; CA-108927.
PA  (HSCR-) HSC RES & DEV LP.
PI  Bull P,  Cox DW,  Thomas G;
DR  WPI; 95-139609/18.
PT  Nucleotide sequence contg. the gene for Wilson disease - used to
PT  treat and detect Wilson disease
PS  Example 3; Fig 11-2; 127pp; English.
CC  The nucleotide sequence of the exon 5/intron borders of the gene encoding
CC  the copper transporting ATPase (ATP7B) associated with Wilson disease
CC  (WD; Q86541). The exon sequence covers the transcript region from
CC  nucleotides 1777 to 1853. The ATP7B coding region covers around 40 kb
CC  of genomic DNA and is divided into 21 exons (see Q86542-Q86549 and
CC  Q86300-Q86309 for other exon/intron borders). The gene can be used to
CC  treat Wilson disease, an autosomal recessive disorder of copper transport
CC  which results in copper accumulation and toxicity to the liver and brain.
```

CC   The DNA sequence can also be used in gene therapy to remove heavy metals
CC   from an organ.
SQ   Sequence   667 BP;   178 A;   162 C;   134 G;   193 T;

alignment_scores:
         Quality: 7.00   Length: 7
           Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x Q86546   ..

Align seg 1/1   to: Q86546   from: 1   to: 667

       136 GlyProThrLeuProAlaGly 142
           |||||||||||||||||||||
       447 GGCCCAACTCTGCCAGCTGGT 467

OM of: US-09-252-691-7056 to: Issued_Patents_NA:*   out_format : pfs

Date: Jun 10, 2000 2:36 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters;
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/USPTO.spool/US09252691/runat_05062000_101737_1274/app_query.fasta.1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oligo.rni
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -DELOP=6.000 -DELEXT=7.000 -START=1
-YGAPEXT=60.000 -YGAPOP=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=human40.cdi -LIST=1000 -DOCALIGN=200
-THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=1000000 -USER=US09252691 -NCPU=6
-ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: Issued_Patents_NA:*
Database sequences: 230463
Database length: 64992525
Search time (sec): 26.990000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-171-389-182	+	7.00	132.17	6.79	46	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-123-936-182	+	7.00	132.17	6.79	46	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-475-228A-182	+	7.00	132.17	6.79	46	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-482-080A-182	+	7.00	132.17	6.79	46	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-12388-182	+	7.00	132.17	6.79	46	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-468-819-73	-	7.00	111.56	95.45	654	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-468-819-75	-	7.00	111.56	95.45	654	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-997-080-40	-	7.00	106.78	176.25	1211	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-997-362-40	-	7.00	106.78	176.25	1211	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-873-970-40	-	7.00	106.78	176.25	1211	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-468-709B-5	-	7.00	106.10	192.33	1322	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-241-664B-5	-	7.00	106.10	192.33	1322	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-03936-5	-	7.00	106.10	192.33	1322	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-389-812-3	+	7.00	105.29	213.47	1468	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-583-318-4	+	7.00	104.47	237.05	1631	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-13	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-19	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-21	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-23	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-25	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-27	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-29	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-31	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-37	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-39	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-41	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-43	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-943-087-45	+	7.00	104.34	241.11	1659	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-780-370A-1	+	7.00	102.06	322.93	2225	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-225-488-6	+	7.00	99.76	433.82	2993	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-751-305-1	+	7.00	98.63	501.18	3460	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-944-087-1	+	7.00	98.51	509.25	3516	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-916-917-1	+	7.00	96.27	678.70	4692	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-972-631-1	+	7.00	96.27	678.70	4692	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-972-629-1	+	7.00	96.27	678.70	4692	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-972-630-1	+	7.00	96.27	678.70	4692	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-672-211-1	+	7.00	96.27	678.70	4692	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-09-225-170-1	+	7.00	96.27	678.70	4692	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-094-948A-4	-	7.00	95.58	741.03	5125	

/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-09319-4	-	7.00	95.58	741.03	5924	51
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-459-586-6	-	7.00	94.46	856.00	5924	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-282-696-6	-	7.00	94.46	856.00	6152	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-557-139-1	-	7.00	94.16	888.79	11703	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-801-263A-8	-	7.00	89.17	1.7e+03	11703	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-09-102-248-8	-	7.00	89.17	1.7e+03	15894	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-905-817-1	-		86.80	2.3e+03	15894	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-741-881-1	-	7.00	86.80	2.4e+03	16656	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-739-158-1	-	7.00	86.43	2.4e+03	16656	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-739-167-1	-	7.00	86.43	2.4e+03	16656	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-404-796-1	-	7.00	86.43	2.4e+03	16656	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-931-869-1	-	7.00	86.43	2.4e+03	16656	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-472-217-1	-	7.00	82.77	3.8e+03	26700	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-488-199-5	-	7.00	82.77	3.8e+03	26700	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-760-534A-1	-	7.00	82.77	3.8e+03	26700	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-09-205-204-15	-	6.00	120.37	30.84	18	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-131-324-5	+	6.00	119.55	34.25	20	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-09-205-204-15	-	6.00	119.55	34.25	20	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-652-127C-4	-	6.00	117.82	42.77	25	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-640-517A-56	+	6.00	117.22	46.17	27	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-640-517A-59	+	6.00	117.22	46.17	27	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-415-823-9	-	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-640-517A-67	+	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-902-623-31	-	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-902-623-34	-	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-086-562-9	-	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-02260-6	-	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-02260-7	-	6.00	116.41	51.28	30	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-292-620A-797	+	6.00	114.99	61.49	36	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-585-684B-1101	+	6.00	114.99	61.49	36	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-585-684B-1493	+	6.00	114.99	61.49	36	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-171-189-183	+	6.00	113.09	78.48	46	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-123-936-183	+	6.00	113.09	78.48	46	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-475-228A-183	+	6.00	113.09	78.48	46	
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-482-080A-183	+	6.00	113.09	78.48	46	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-12388-183	+	6.00	113.09	78.48	53	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-316-293-56	+	6.00	111.99	90.36	57	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-089-974B-31	+	6.00	111.42	97.15	71	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-434-001-173	+	6.00	109.72	120.89	71	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-433-585-173	+	6.00	109.72	120.89	71	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-434-425-173	+	6.00	109.72	120.89	71	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-437-667-173	+	6.00	109.72	120.89	71	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US93-02260-7	-	6.00	109.72	120.89	73	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-906-955-173	+	6.00	109.50	124.28	7	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-06060-173	-	6.00	109.72	120.89		
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:US-08-945-734-12	+	6.00	109.50	124.28	79	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-09451-12	-	6.00	108.89	134.45	79	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-447-169A-95	-	6.00	108.89	134.45	92	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-233-012C-95	-	6.00	107.71	156.46	113	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-089-974B-19	+	6.00	106.11	192.00	113	
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-710-082-8	+	6.00	106.11	192.00	116	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-710-082-17	+	6.00	105.91	197.07	116	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-487-867-9	+	6.00	105.91	197.07	116	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-487-867-10	+	6.00	105.91	197.07	156	
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-487-867-11	+	6.00	105.91	197.07	156	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-09358-9	+	6.00	103.61	264.68	166	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-09358-10	+	6.00	103.61	264.68	177	
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US96-09358-11	+	6.00	103.13	281.56	177	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-171-383A-8	-	6.00	102.63	300.13	177	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-461-041-8	-	6.00	102.63	300.13	177	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-089-974B-15	-	6.00	102.63	300.13	177	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-256-568B-58	-	6.00	102.63	300.13	177	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-256-568B-75	-	6.00	102.63	300.13	206	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-256-568B-76	-	6.00	101.45	349.07	264	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-256-568B-77	-	6.00	99.53	446.85	274	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-256-568B-78	-	6.00	99.24	463.69	284	
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-866-340-15	-	6.00	98.96	480.54	346	
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-182-961B-1	-	6.00	97.43	584.92		
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-765-332-101	+	6.00				
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-902-623-42	+	6.00				
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-967-101-33	+	6.00				

Sequence	strand		pct	value	index
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-592-541-33	+	6.00	97.43	584.92	346
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-171-383A-7	-	6.00	96.34	672.40	398
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-461-041-7	-	6.00	96.34	672.40	398
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-486-013-61	-	6.00	95.94	707.72	419
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-482-279-61	-	6.00	95.94	707.72	419
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-342-268-61	-	6.00	95.94	707.72	419
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-826-532-1	!	6.00	95.78	722.85	428
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-946-914-35	+	6.00	95.40	758.15	449
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-387-942C-9	+	6.00	95.23	774.96	459
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-623-909A-11	+	6.00	95.22	776.64	460
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-602-264A-13	+	6.00	94.90	808.57	479
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-215-805A-81	+	6.00	94.65	835.46	495
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-532-828B-13	+	6.00	94.32	870.74	516
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-532-828B-14	+	6.00	94.32	870.74	516
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-484-956-70	+	6.00	94.06	900.97	534
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-757-653-70	+	6.00	94.06	900.97	534
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-956-69	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-956-71	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-956-72	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-956-82	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-956-83	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-757-653-69	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-757-653-71	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-757-653-72	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-757-653-82	+	6.00	94.03	904.33	536
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-757-653-83	+	6.00	93.76	936.24	555
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-920-440B-1	+	6.00	93.76	936.24	555
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12779-1	+	6.00	93.70	942.96	559
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-15781-1	+	6.00	93.30	993.33	589
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-948-197-2	+	6.00	93.23	1.0e+03	594
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-454-196-3	-	6.00	93.15	1.0e+03	600
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-652-127C-5	-	6.00	92.79	1.1e+03	629
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-09430-15	-	6.00	92.56	1.1e+03	648
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-670-186-1	+	6.00	92.56	1.1e+03	648
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-225-989-22	-	6.00	92.56	1.1e+03	648
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-570-923-22	-	6.00	92.41	1.1e+03	660
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-580-014-22	-	6.00	91.90	1.2e+03	705
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-325-553-28	+	6.00	91.90	1.2e+03	705
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-394-152A-28	+	6.00	91.82	1.2e+03	713
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-225-989-7	-	6.00	91.82	1.2e+03	713
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-570-923-7	-	6.00	91.78	1.2e+03	716
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-580-014-7	-	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-238-027-3	-	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-385-335A-10	+	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-09-020-466-3	+	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-698-551-5	-	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-602-228-5	-	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12724-5	-	6.00	91.72	1.2e+03	722
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-05083C-178	-	6.00	91.63	1.2e+03	730
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-06846A-77	!	6.00	91.63	1.2e+03	730
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-676-279-42	-	6.00	91.23	1.3e+03	769
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-445-909A-28	+	6.00	91.12	1.3e+03	780
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-494-440B-5	-	6.00	90.50	1.4e+03	845
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-266-451B-1	-	6.00	90.09	1.5e+03	891
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-249-322A-77	-	6.00	90.09	1.5e+03	891
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-734-591A-77	-	6.00	90.06	1.5e+03	894
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-469-660-182	!	6.00	89.64	1.6e+03	944
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-839-031A-5	+	6.00	91.63	1.2e+03	730
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-469-526A-5	+	6.00	91.63	1.2e+03	730
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-149-097D-27	-	6.00	89.53	1.6e+03	957
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-599-480-1	-	6.00	89.53	1.6e+03	957
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-842-199-1	-	6.00	89.53	1.6e+03	957
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-00190-1	+	6.00	89.18	1.7e+03	1001
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-728-259A-10	+	6.00	89.18	1.7e+03	1001
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-473-486-10	+	6.00	89.16	1.7e+03	1004
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-841-646-10	-	6.00	89.16	1.7e+03	1004
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-147-023-10	-	6.00	89.16	1.7e+03	1004
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-147-570-10	-	6.00	89.16	1.7e+03	1004
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-449-700-10	-	6.00	89.16	1.7e+03	1004
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-449-699A-10	-	6.00	89.16	1.7e+03	1004
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-114-461-1	+	6.00	89.04	1.7e+03	1020
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-192-156-1	+	6.00	89.04	1.7e+03	1020
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-370-789-1	+	6.00	89.04	1.7e+03	1020
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-698-550-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-602-228-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-649-341A-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-494-440B-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-533-901B-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-839-032A-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-839-031A-7	-	6.00	89.01	1.7e+03	1023
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US92-01196-5	-		89.01	1.7e+03	10
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-12724-7	-		89.01	1.7e+03	10
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-743-637B-191	+	6.00	88.97	1.7e+03	102
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-815-667-11	+	6.00	88.74	1.8e+03	1059
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-385-335A-13	+	6.00	88.73	1.8e+03	1061
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-537-434-2	-	6.00	88.55	1.8e+03	1086
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-886-633-1	+	6.00	88.45	1.8e+03	1100
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-933-750C-77	+	6.00	88.40	1.9e+03	1107
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-523-855A-2	-	6.00	88.30	1.9e+03	1121
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-613-965-1	+	6.00	88.24	1.9e+03	1130
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-918-966-1	+	6.00	88.24	1.9e+03	1130
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-921-655-1	!		88.24	1.9e+03	1130
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-454-196-1	+	6.00	88.17	1.9e+03	1140
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-457-245-4	-	6.00	88.07	1.9e+03	1155
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-869-933-10	+	6.00	87.94	2.0e+03	1174
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-107-576-26	+	6.00	87.92	2.0e+03	1178
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-756-387B-1	+	6.00	87.79	2.0e+03	1198
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-756-387B-3	+	6.00	87.79	2.0e+03	1198
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-107-676-2	+	6.00	87.70	2.0e+03	1211
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-846-762-100	+	6.00	87.67	2.0e+03	1217
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-031-485-17	+	6.00	87.60	2.1e+03	1227
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-031-485-18	+	6.00	87.60	2.1e+03	1227
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-847-429A-17	+	6.00	87.60	2.1e+03	1227
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-847-429A-18	+	6.00	87.60	2.1e+03	1227
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-031-485-14	+	6.00	87.60	2.1e+03	1228
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-031-485-16	+	6.00	87.60	2.1e+03	1228
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-847-429A-14	+	6.00	87.60	2.1e+03	1228
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-847-429A-16	+	6.00	87.60	2.1e+03	1228
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-539-798-1	-	6.00	87.41	2.1e+03	1258
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-329-560-1	-	6.00	87.41	2.1e+03	1258
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-764-731B-9	-	6.00	87.40	2.1e+03	1259
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-672-814D-12	+	6.00	87.40	2.1e+03	1263
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-107-576-26	+	6.00	87.38	2.1e+03	1263
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-532-828B-12	+	6.00	87.32	2.1e+03	1272
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-068-109-1	-	6.00	87.32	2.2e+03	1277
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-841-646-14	-	6.00	87.29	2.2e+03	1277
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-147-023-14	-	6.00	87.29	2.2e+03	1277
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-447-570-14	-	6.00	87.29	2.2e+03	1277
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-449-699A-14	-	6.00	87.29	2.2e+03	1287
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-476-008-41	-	6.00	87.23	2.2e+03	1287
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-306-063-41	-	6.00	87.23	2.2e+03	1287
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-833-485-41	-	6.00	87.23	2.2e+03	1287
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-445-909A-16	+	6.00	87.19	2.2e+03	1294
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-865-368-1	+	6.00	87.19	2.2e+03	1295
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-256-368-2	!	6.00	86.63	2.3e+03	1391
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-865-336-2	-	6.00	86.62	2.3e+03	1392
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-478-097A-37	-	6.00	86.60	2.3e+03	1396
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-151-611-2	+	6.00	86.57	2.4e+03	1401
/cgn2_6/ptodata/1/ina/backfiles1.seq:5171840-8	+	6.00	86.56	2.4e+03	1404	! P
/cgn2_6/ptodata/1/ina/backfiles1.seq:5480796-8	+		86.56	2.4e+03	1404	! P
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-07127A-3	+		86.41	2.4e+03	1
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-845-742-1	-	6.00	86.37	2.4e+03	1438
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-539-756-3	-	6.00	86.32	2.4e+03	1448
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-377-292-7	-	6.00	86.32	2.4e+03	1448
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-989-847-5	-	6.00	86.32	2.4e+03	1448
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-622-354-2	-	6.00	86.19	2.5e+03	1472

```
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-554-612C-51 +    6.00  80.93  4.8e+03 2898 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-027-337-9 -      6.00  80.93  4.8e+03 2900 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-00419-4 -  6.00  80.92  4.8e+03 2901 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-554-612C-10 +    6.00  80.90  4.8e+03 2909 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-554-612C-11 +    6.00  80.90  4.8e+03 2909 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-500-857A-1 +     6.00  80.89  4.9e+03 2915 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-357-073-1 -      6.00  80.86  4.9e+03 2924 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-281-193-1 +      6.00  80.83  4.9e+03 2935 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-422-716-1 +      6.00  80.83  4.9e+03 2935 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-735-716-1 +      6.00  80.83  4.9e+03 2935 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-555-568B-1 +     6.00  80.83  4.9e+03 2935 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-08069-1 +  6.00  80.82  4.9e+03 2939 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-01119-512-2 +  6.00  80.82  4.9e+03 2939 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-488-015B-2 +     6.00  80.69  5.0e+03 2990 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-671-817A-1 +     6.00  80.62  5.0e+03 3016 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-318-826A-7 +     6.00  80.62  5.0e+03 3016 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-370-156-5 +      6.00  80.62  5.0e+03 3016 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-814-095-5 +       6.00  80.44  5.1e+03 3088 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-418-444A-1 +     6.00  80.42  5.2e+03 3096 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-318-826A-6 +     6.00  80.42  5.2e+03 3096 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-370-156-3 +      6.00  80.42  5.2e+03 3096 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-814-095-3 +       6.00  80.42  5.2e+03 3096 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-162-809-1 +      6.00  80.33  5.2e+03 3133 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-693-174-5 +      6.00  80.27  5.3e+03 3157 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-500-857A-9 +     6.00  80.21  5.3e+03 3258 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-729-955A-2 +     6.00  80.02  5.4e+03 3283 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-815-667-14 +     6.00  79.96  5.5e+03 3292 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-611-107-33 +      6.00  79.94  5.5e+03 3318 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-326-670A-1 -     6.00  79.88  5.5e+03 3336 ! Pate
/cgn2_6/ptodata/1/ina/backfiles1.seq:457037-1 -          6.00  79.84  5.5e+03 3336 ! Pate
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-469-803B-1 +     6.00  79.77  5.6e+03 3366 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-267-803B-1 +     6.00  79.77  5.6e+03 3366 !
/cgn2_6/ptodata/1/ina/backfiles1.seq:5190871-1 -         6.00  79.69  5.7e+03 3401 ! Pate
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-631-717D-1 -     6.00  79.58  5.7e+03 3447 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-313-185-57 +     6.00  79.55  5.7e+03 3461 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-389-564B-1 +     6.00  79.55  5.8e+03 3461 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-466-047B-1 +     6.00  79.53  5.8e+03 3472 !
/cgn2_6/ptodata/1/ina/backfiles1.seq:5244792-1 -         6.00  79.52  5.8e+03 3472 ! Pate
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-960-389-1 +      6.00  79.52  5.8e+03 3475 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-448-170-5 -      6.00  79.46  5.8e+03 3501 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-166-717D-1 -     6.00  79.46  5.8e+03 3503 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-658-665-64 -     6.00  79.42  5.8e+03 3521 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-828-788A-9 -     6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-040-751-4 -      6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-349-867-24 -     6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-349-867-28 -     6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-239-476-24 -     6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-239-476-28 -     6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-356-034-7 -      6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-291-368-1 -      6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US92-11337-9 -  6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-05431-24 - 6.00  79.42  5.9e+03 3522 ! Pate
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-05431-28 - 6.00  79.42  5.9e+03 3522 ! Pate
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-10310-1 -  6.00  79.42  5.9e+03 3522 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-598-305A-24 -    6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-598-305A-28 -    6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-639-923A-24 -    6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-639-923A-28 -    6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-754-490-9 -       6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-754-490-11 -      6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-962-190-1 -      6.00  79.40  5.9e+03 3531 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US92-11337-7 -  6.00  79.40  5.9e+03 3531 ! Pate
/cgn2_6/ptodata/1/ina/backfiles1.seq:5169629-1 -         6.00  79.40  5.9e+03 3532 ! Pate
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-09-205-204-1 -      6.00  79.40  5.9e+03 3532 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-476-062A-40 -    6.00  79.39  5.9e+03 3533 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-01314-40 - 6.00  79.39  5.9e+03 3533 !
```

```
/cgn2_6/ptodata/1/ina/backfiles1.seq:5424399-1 -         6.00  79.39  5.9e+03 3533 ! P
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-841-178-24 -     6.00  79.39  5.9e+03 3534 -
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-841-178-25 -     6.00  79.39  5.9e+03 3534 -
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-841-178-26 -     6.00  79.39  5.9e+03 3534 -
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-754-490-25 -      6.00  79.39  5.9e+03 3534 -
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-754-490-27 -      6.00  79.38  5.9e+03 3537 -
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-566-398-38 -     6.00  79.38  5.9e+03 3537 -
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-162-809-9 -      6.00  79.36  5.9e+03 3546 -
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-602-737-15 -     6.00  79.34  5.9e+03 3558 -
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-097-997A-12 +    6.00  79.33  5.9e+03 3561 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:US-08-306-691B-44 + 6.00  79.32  5.9e+03 3565 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-06251-20 + 6.00  79.32  5.9e+03 3567 3
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-828-788A-15 -    6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-356-034-5 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-602-737-1 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-1 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-3 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-5 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-7 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-9 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-11 -     6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-58 -     6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-980-071-60 -     6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-757-536-1 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-757-536-5 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-757-536-7 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-757-536-9 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-757-536-11 -     6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-1 -       6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-3 -       6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-5 -       6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-9 -       6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-11 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-58 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-60 -      6.00  79.32  5.9e+03 3567 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US92-11337-5 -  6.00  79.32  5.9e+03 3567 3
/cgn2_6/ptodata/1/ina/backfiles1.seq:5188960-5 -         6.00  79.32  5.9e+03 3567 ! P
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-602-737-7 -      6.00  79.29  5.9e+03 3579 -
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-754-490-29 -      6.00  79.29  5.9e+03 3579 -
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-162-809-13 -     6.00  79.27  6.0e+03 3591 -
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-951-715A-6 -     6.00  79.20  6.0e+03 3624 -
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-459-448A-6 -     6.00  79.20  6.0e+03 3624 -
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-459-595A-6 -     6.00  79.20  6.0e+03 3624 -
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-314-093-58 -      6.00  79.16  6.0e+03 3639 +
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-737-524B-26 +    6.00  79.16  6.1e+03 3642 -
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-946-026-16 -      6.00  79.04  6.1e+03 3699 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-680-336-29 +     6.00  79.02  6.4e+03 3708 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-215-805A-1 +     6.00  78.73  6.4e+03 3848 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-862-021B-9 +     6.00  78.69  6.4e+03 3870 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-313-288B-9 -     6.00  78.69  6.4e+03 3870 +
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03164-9 -  6.00  78.64  6.5e+03 3893 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-09422-18 - 6.00  78.64  6.5e+03 3893 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-138-133-1 +      6.00  78.64  6.5e+03 3893 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-138-641-3 +      6.00  78.58  6.5e+03 3923 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-138-133-3 +      6.00  78.56  6.5e+03 3934 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-465-609-4 -      6.00  78.54  6.6e+03 3946 !
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-226-568-18 -      6.00  78.54  6.6e+03 3946 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-077-848A-1 -     6.00  78.54  6.6e+03 3946 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:US-09-211-640-1 -   6.00  78.54  6.6e+03 4003 39
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:US-08-257-999-1 -   6.00  78.42  6.6e+03 4029 40
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-02560-1 -  6.00  78.39  6.7e+03 4029 -
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-862-021B-9 +     6.00  78.37  6.7e+03 4029 -
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-313-288B-9 -     6.00  78.37  6.7e+03 3870 40
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-09422-18 - 6.00  78.25  6.8e+03 4
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-434-823-1 -      6.00  78.23  6.8e+03 4106 -
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-366-1 -      6.00  78.23  6.8e+03 4106 -
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-702-572-14 -     6.00  78.23  6.8e+03 4106 -
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-204-675-1 -      6.00  78.12  6.9e+03 4164 -
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-660-754-1 -      6.00  78.12  6.9e+03 4164 -
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-796-364-1 -      6.00  78.12  6.9e+03 4164 -
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-02520-1 -  6.00  78.12  6.9e+03 4164 41
```

```
/cgn2_6/ptodata/1/ina/5x_COMB.seq:US:US-08-xxx-xxx-n  [+/-]  6.00  75.xx  9.xe+03  57xx
```

```
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-944-868A-11 +        5.00                    24 !
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-944-423A-11 +        5.00  99.06  474.31     24 !   24
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US94-10957-1 -            99.06  474.31        !
/cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-182-961B-17 +      5.00    99.06  474.31     25 !
/cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-623-891-64 -       5.00    98.74  493.98     25 !   25
/cgn2_6/ptodata/1/1na/5C_COMB.seq:US-08-692-787-52 -       5.00    98.74  493.98     25 !
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-713-557B-29 -      5.00    98.74  493.98     25 !
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-668-650B-18 -      5.00    98.74  493.98     25 !
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-859-998-329 +      5.00    98.74  493.98     25 !
/cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-859-998-799 +      5.00    98.74  493.98     25 !
/cgn2_6/ptodata/1/1na/6_COMB.seq:US-08-922-955-1 +         5.00    98.74  493.98     25 !   25
/cgn2_6/ptodata/1/1na/PCTUS_COMB.seq:PCT-US94-07659-19 +   5.00    98.74  493.98        ! Seq
                                                                                          25
```

seq_documentation_block: ◀
; Sequence 182, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
;   APPLICANT: Edwards, Cynthia A.
;   APPLICANT: Cantor, Charles R.
;   APPLICANT: Andrews, Beth M.
;   APPLICANT: Turin, Lisa M.
;   APPLICANT: Fry, Kirk E.
;   TITLE OF INVENTION: Sequence-Directed DNA Binding
;   TITLE OF INVENTION: Molecules, Compositions and Methods
;   NUMBER OF SEQUENCES: 641
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genelabs Technologies, Inc.
;   STREET: 505 Penobscot Drive
;   CITY: Redwood City
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94063
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/171,389
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/123,936
;   FILING DATE: 17-SEP-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/996,783
;   FILING DATE: 23-DEC-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/723,618
;   FILING DATE: 27-JUN-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/081,070
;   FILING DATE: 22-JUN-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fabian, Gary R.
;   REGISTRATION NUMBER: 33,875
;   REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 324-0880
;   TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 46 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   INDIVIDUAL ISOLATE: Human hsp70B gene
US-08-171-389-182

alignment_scores:
           Quality: 7.00   Length: 7
           Ratio: 1.000   Gaps: 0
  Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-171-389-182 ..

Align seg 1/1 to: US-08-171-389-182   from: 1   to: 46

      98 AlaArgLeuThrGlnProGly 104
         |||||||||||||||||||||
       3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: /cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-123-936-182

seq_documentation_block:
; Sequence 182, Application US/08123936
; Patent No. 5726014
; GENERAL INFORMATION:
;   APPLICANT: Edwards, Cynthia A.
;   APPLICANT: Cantor, Charles R.
;   APPLICANT: Andrews, Beth M.
;   APPLICANT: Turin, Lisa M.
;   TITLE OF INVENTION: Screening Assay for the Detection of
;   TITLE OF INVENTION: DNA-Binding Molecules
;   NUMBER OF SEQUENCES: 640
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genelabs Technologies, Inc.
;   STREET: 505 Penobscot Drive
;   CITY: Redwood City
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94063
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/123,936
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/996,783
;   FILING DATE: 23-DEC-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/723,618
;   FILING DATE: 27-JUN-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fabian, Gary R.
;   REGISTRATION NUMBER: 33,875
;   REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 324-0880
;   TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 46 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   INDIVIDUAL ISOLATE: Human hsp70B gene
US-08-123-936-182

alignment_scores:
           Quality: 7.00   Length: 7

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 11:25:46 ; Search time 12.47 Seconds
(without alignments)
257.025 Million cell updates/sec

Title: US-09-252-691-7056
Perfect score: 222
Sequence: 1 AIMRQLITPENTMTKTSFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 145308 seqs, 14437401 residues

Word size : 0

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	3.2	416	2	US-08-225-488-7	Sequence 7, Appl1
2	6	2.7	22	1	US-08-769-309A-21	Sequence 21, Appl1
3	6	2.7	99	1	US-08-449-639-3	Sequence 3, Appl1
4	6	2.7	93	2	US-08-902-623-43	Sequence 43, Appl1
5	6	2.7	165	1	US-08-215-805A-82	Sequence 82, Appl1
6	6	2.7	179	2	US-08-933-750C-28	Sequence 28, Appl1
7	6	2.7	242	2	US-08-622-352A-3	Sequence 3, Appl1
8	6	2.7	292	1	US-08-769-309A-3	Sequence 3, Appl1
9	6	2.7	327	1	US-08-375-962B-12	Sequence 12, Appl1
10	6	2.7	327	2	US-08-562-114B-12	Sequence 12, Appl1
11	6	2.7	339	2	US-08-855-714-3	Sequence 3, Appl1
12	6	2.7	365	2	US-08-467-559B-9	Sequence 9, Appl1
13	6	2.7	403	2	US-08-846-762-89	Sequence 89, Appl1
14	6	2.7	409	2	US-09-031-485-15	Sequence 15, Appl1
15	6	2.7	409	3	US-08-847-429A-15	Sequence 15, Appl1
16	6	2.7	446	3	US-08-956-254-2	Sequence 2, Appl1
17	6	2.7	448	2	US-09-015-815-1	Sequence 1, Appl1
18	6	2.7	500	2	US-08-987-519-1	Sequence 1, Appl1
19	6	2.7	500	2	US-08-987-519-2	Sequence 2, Appl1
20	6	2.7	501	2	US-08-987-519-3	Sequence 3, Appl1
21	6	2.7	507	2	US-08-409-122-2	Sequence 2, Appl1
22	6	2.7	507	2	US-08-408-669-2	Sequence 2, Appl1
23	6	2.7	516	2	US-09-019-201A-3	Sequence 3, Appl1
24	6	2.7	532	3	US-08-737-336-6	Sequence 6, Appl1
25	6	2.7	586	2	US-08-630-822A-70	Sequence 70, Appl1
26	6	2.7	586	2	US-09-005-069-70	Sequence 70, Appl1
27	6	2.7	597	1	US-08-374-155A-12	Sequence 12, Appl1
28	6	2.7	597	2	US-08-785-396-12	Sequence 12, Appl1
29	6	2.7	600	1	US-08-374-155A-10	Sequence 10, Appl1
30	6	2.7	600	2	US-08-785-396-10	Sequence 10, Appl1
31	6	2.7	614	1	US-08-543-881-2	Sequence 2, Appl1
32	6	2.7	614	1	US-08-291-299-2	Sequence 2, Appl1
33	6	2.7	614	4	PCT-US94-00119-2	Sequence 2, Appl1
34	6	2.7	614	1	PCT-US95-10579-2	Sequence 4, Appl1
35	6	2.7	629	2	US-08-374-155A-4	Sequence 4, Appl1
36	6	2.7	629	2	US-08-785-396-4	Sequence 4, Appl1
37	6	2.7	661	2	US-08-417-174-27	Sequence 27, Appl1
38	6	2.7	661	2	US-08-417-174-121	Sequence 121, App
39	6	2.7	661	2	US-08-525-742-2	Sequence 2, Appl1
40	6	2.7	661	2	US-08-231-565A-27	Sequence 27, Appl1
41	6	2.7	661	2	US-09-007-961-27	Sequence 27, Appl1
42	6	2.7	668	2	US-07-891-942G-6	Sequence 6, Appl1
43	6	2.7	879	1	US-05-554-612C-1	Sequence 1, Appl1
44	6	2.7	897	2	US-07-960-389-2	Sequence 2, Appl1
45	6	2.7	947	2	US-08-500-857A-4	Sequence 4, Appl1
46	6	2.7	953	1	US-08-500-857A-2	Sequence 2, Appl1
47	6	2.7	955	2	US-08-500-857A-10	Sequence 10, Appl1
48	6	2.7	980	1	US-08-220-151-5	Sequence 5, Appl1
49	6	2.7	980	2	US-08-413-118-5	Sequence 5, Appl1
50	6	2.7	980	3	US-08-473-446-5	Sequence 5, Appl1
51	6	2.7	1148	1	US-08-313-185-58	Sequence 58, Appl1
52	6	2.7	1243	2	US-08-557-139-2	Sequence 2, Appl1
53	6	2.7	1290	1	US-08-138-641-2	Sequence 2, Appl1
54	6	2.7	1290	2	US-08-138-133-2	Sequence 2, Appl1
55	6	2.7	1350	2	US-08-319-866-9	Sequence 9, Appl1
56	6	2.7	1745	2	US-09-031-485-33	Sequence 33, Appl1
57	6	2.7	1745	2	US-08-847-429A-33	Sequence 33, Appl1
58	6	2.7	1780	1	US-08-769-309A-5	Sequence 5, Appl1
59	6	2.7	2206	2	US-07-852-260-2	Sequence 2, Appl1
60	6	2.7	2206	1	US-08-461-503-2	Sequence 2, Appl1
61	6	2.7	2938	4	PCT-US94-00198-3	Sequence 3, Appl1
62	6	2.7	3011	1	US-08-188-281B-1	Sequence 1, Appl1
63	6	2.7	3011	2	US-08-453-552-1	Sequence 1, Appl1
64	6	2.7	3011	1	US-08-710-637-1	Sequence 1, Appl1
65	6	2.7	3011	4	PCT-US93-00907-1	Sequence 1, Appl1
66	6	2.7	3011	4	PCT-US94-07280-1	Sequence 1, Appl1
67	6	2.7	3011	4	PCT-US95-01087-1	Sequence 1, Appl1
68	6	2.7	3144	1	US-08-246-982A-6	Sequence 6, Appl1
69	6	2.7	3144	2	US-08-453-265-6	Sequence 6, Appl1
70	6	2.7	3144	2	US-08-457-273B-42	Sequence 42, Appl1
71	6	2.7	3144	2	US-07-642-734C-4	Sequence 4, Appl1
72	5	2.3	3567	2	US-08-067-387-18	Sequence 18, Appl1
73	5	2.3	7	4	PCT-US94-05796-18	Sequence 18, Appl1
74	5	2.3	9	2	US-09-070-756-6	Sequence 6, Appl1
75	5	2.3	10	2	US-08-556-597-170	Sequence 170, App
76	5	2.3	11	1	US-08-751-767A-73	Sequence 73, Appl1
77	5	2.3	11	2	US-08-156-552A-12	Sequence 12, Appl1
78	5	2.3	12	2	US-08-260-582-36	Sequence 36, Appl1
79	5	2.3	12	2	US-08-764-640-189	Sequence 189, App
80	5	2.3	12	4	PCT-US95-05471-36	Sequence 36, App
81	5	2.3	13	2	US-08-764-640-173	Sequence 173, App
82	5	2.3	13	2	US-08-764-640-179	Sequence 179, App
83	5	2.3	13	2	US-08-764-640-180	Sequence 180, App
84	5	2.3	13	2	US-08-764-640-183	Sequence 183, App
85	5	2.3	13	2	US-08-764-640-186	Sequence 186, App
86	5	2.3	13	2	US-08-764-640-228	Sequence 228, App
87	5	2.3	13	2	US-08-764-640-230	Sequence 230, App
88	5	2.3	14	1	US-08-463-620-6	Sequence 6, Appl1
89	5	2.3	14	1	US-08-463-620-7	Sequence 7, Appl1
90	5	2.3	14	2	US-08-764-640-12	Sequence 12, Appl1
91	5	2.3	14	2	US-08-764-640-38	Sequence 38, Appl1
92	5	2.3	14	2	US-08-764-640-52	Sequence 52, Appl1
93	5	2.3	14	2	US-08-764-640-172	Sequence 172, App
94	5	2.3	14	2	US-08-764-640-174	Sequence 174, App
95	5	2.3	14	2	US-08-764-640-175	Sequence 175, App
96	5	2.3	14	2	US-08-764-640-176	Sequence 176, App
97	5	2.3	14	2	US-08-764-640-177	Sequence 177, App
98	5	2.3	14	2	US-08-764-640-178	Sequence 178, App
99	5	2.3	14	2	US-08-764-640-181	Sequence 181, App
100	5	2.3	14	2	US-08-764-640-182	Sequence 182, App
101	5	2.3	14	2	US-08-764-640-184	Sequence 184, App

Row			n1	n2	Patent Number	Sequence
102	5	2.3	14	2	US-08-764-640-229	Sequence 229, App
103	5	2.3	14	2	US-08-934-222-23	Sequence 23, Appl
104	5	2.3	14	2	US-08-933-402-23	Sequence 23, Appl
105	5	2.3	14	2	US-09-207-621-23	Sequence 23, Appl
106	5	2.3	14	2	US-08-224-917-6	Sequence 6, Appl
107	5	2.3	14	2	US-08-224-917-7	Sequence 23, Appl
108	5	2.3	14	2	US-08-532-818-23	Sequence 6, Appl
109	5	2.3	14	2	US-08-914-853-6	Sequence 7, Appl
110	5	2.3	14	2	US-08-914-853-7	Sequence 6, Appl
111	5	2.3	14	4	PCT-US95-03934A-6	Sequence 7, Appl
112	5	2.3	14	4	PCT-US95-03934A-7	Sequence 17, Appl
113	5	2.3	15	1	US-08-268-251-17	Sequence 95, Appl
114	5	2.3	15	2	US-08-637-759B-95	Sequence 5, Appl
115	5	2.3	15	5	US-09-049-577-5	Sequence 95, Appl
116	5	2.3	15	2	PCT-US93-01112-17	Sequence 17, Appl
117	5	2.3	15	5	5227466-5	Patent No. 5227466
118	5	2.3	16	2	US-08-291-349A-3	Sequence 3, Appl
119	5	2.3	16	2	US-09-025-706-8	Sequence 8, Appl
120	5	2.3	16	4	US-08-394-748A-11	Sequence 11, Appl
121	5	2.3	17	4	PCT-US95-02478-11	Sequence 11, Appl
122	5	2.3	17	4	US-08-480-190-123	Sequence 123, App
123	5	2.3	17	4	US-08-488-379-123	Sequence 123, App
124	5	2.3	18	4	PCT-US93-07545-123	Sequence 123, App
125	5	2.3	18	1	US-08-257-528B-50	Sequence 50, Appl
126	5	2.3	18	1	US-08-460-602A-50	Sequence 50, Appl
127	5	2.3	18	2	US-08-463-966A-50	Sequence 50, Appl
128	5	2.3	18	2	US-08-465-217A-50	Sequence 50, Appl
129	5	2.3	18	2	US-08-464-329A-50	Sequence 50, Appl
130	5	2.3	18	2	US-08-764-640-6	Sequence 6, Appl
131	5	2.3	18	2	US-08-764-640-44	Sequence 44, Appl
132	5	2.3	18	2	US-08-764-640-53	Sequence 53, Appl
133	5	2.3	18	2	US-08-764-640-58	Sequence 58, Appl
134	5	2.3	18	2	US-08-462-507A-50	Sequence 50, Appl
135	5	2.3	19	2	US-08-467-881A-50	Sequence 7, Appl
136	5	2.3	19	2	US-08-764-640-7	Sequence 8, Appl
137	5	2.3	19	2	US-08-764-640-8	Sequence 48, Appl
138	5	2.3	19	2	US-08-764-640-48	Sequence 56, Appl
139	5	2.3	19	3	US-08-764-640-56	Sequence 164, App
140	5	2.3	19	2	US-08-764-640-164	Sequence 21, Appl
141	5	2.3	20	2	US-07-678-974D-23	Sequence 23, Appl
142	5	2.3	20	3	US-08-726-306A-31	Sequence 31, Appl
143	5	2.3	20	2	US-08-945-168-28	Sequence 28, Appl
144	5	2.3	21	2	US-08-615-942A-7	Sequence 7, Appl
145	5	2.3	21	3	US-08-726-306A-15	Sequence 15, Appl
146	5	2.3	22	1	US-07-828-450-15	Sequence 15, Appl
147	5	2.3	22	2	US-08-196-630A-8	Sequence 8, Appl
148	5	2.3	23	1	US-08-176-500-89	Sequence 89, Appl
149	5	2.3	23	2	US-08-268-251-47	Sequence 47, Appl
150	5	2.3	23	1	US-08-471-052A-89	Sequence 89, Appl
151	5	2.3	23	2	US-08-189-331-89	Sequence 89, Appl
152	5	2.3	23	1	US-08-471-939-89	Sequence 89, Appl
153	5	2.3	23	1	US-08-471-800-89	Sequence 89, Appl
154	5	2.3	23	2	US-08-471-068-89	Sequence 89, Appl
155	5	2.3	23	1	PCT-US93-01112-47	Sequence 47, Appl
156	5	2.3	25	1	US-08-764-640-231	Sequence 231, App
157	5	2.3	26	2	US-08-453-289-1	Sequence 1, Appl
158	5	2.3	26	4	US-07-942-245-258	Sequence 258, App
159	5	2.3	26	1	US-08-574-086-1	Sequence 1, Appl
160	5	2.3	27	2	US-08-559-524A-13	Sequence 13, Appl
161	5	2.3	27	2	US-07-668-648-12	Sequence 12, Appl
162	5	2.3	27	3	US-08-429-998-12	Sequence 12, Appl
163	5	2.3	32	2	US-08-431-333-12	Sequence 12, Appl
164	5	2.3	33	4	PCT-US91-02321-12	Sequence 12, Appl
165	5	2.3	33	1	US-08-606-143-38	Sequence 38, Appl
166	5	2.3	33	2	US-08-257-528B-41	Sequence 41, Appl
167	5	2.3	33	2	US-08-257-528B-80	Sequence 80, Appl
168	5	2.3	33	1	US-08-257-528B-81	Sequence 81, Appl
169	5	2.3	33	1	US-08-257-528B-83	Sequence 83, Appl
170	5	2.3	33	2	US-08-257-528B-84	Sequence 84, Appl
171	5	2.3	33	1	US-08-460-602A-41	Sequence 41, Appl
172	5	2.3	33	1	US-08-460-602A-80	Sequence 80, Appl
175	5	2.3	33	1	US-08-460-602A-81	Sequence 81, Appl
176	5	2.3	33	1	US-08-460-602A-83	Sequence 83, Appl
177	5	2.3	33	1	US-08-460-602A-84	Sequence 84, Appl
178	5	2.3	33	1	US-08-463-966A-41	Sequence 41, Appl
179	5	2.3	33	1	US-08-463-966A-80	Sequence 80, Appl
180	5	2.3	33	1	US-08-463-966A-81	Sequence 81, Appl
181	5	2.3	33	1	US-08-463-966A-83	Sequence 83, Appl
182	5	2.3	33	1	US-08-463-966A-84	Sequence 84, Appl
183	5	2.3	33	1	US-08-465-217A-41	Sequence 41, Appl
184	5	2.3	33	1	US-08-465-217A-80	Sequence 80, Appl
185	5	2.3	33	1	US-08-465-217A-81	Sequence 81, Appl
186	5	2.3	33	1	US-08-465-217A-83	Sequence 83, Appl
187	5	2.3	33	1	US-08-465-217A-84	Sequence 84, Appl
188	5	2.3	33	2	US-08-464-329A-41	Sequence 41, Appl
189	5	2.3	33	2	US-08-464-329A-80	Sequence 80, Appl
190	5	2.3	33	2	US-08-464-329A-81	Sequence 81, Appl
191	5	2.3	33	2	US-08-464-329A-83	Sequence 83, Appl
192	5	2.3	33	2	US-08-464-329A-84	Sequence 84, Appl
193	5	2.3	33	2	US-08-462-507A-41	Sequence 41, Appl
194	5	2.3	33	2	US-08-462-507A-80	Sequence 80, Appl
195	5	2.3	33	2	US-08-462-507A-81	Sequence 81, Appl
196	5	2.3	33	2	US-08-462-507A-83	Sequence 83, Appl
197	5	2.3	33	2	US-08-462-507A-84	Sequence 84, Appl
198	5	2.3	33	2	US-08-467-881A-41	Sequence 41, Appl
199	5	2.3	33	2	US-08-467-881A-80	Sequence 80, Appl
200	5	2.3	33	2	US-08-467-881A-81	Sequence 81, Appl
201	5	2.3	33	2	US-08-467-881A-83	Sequence 83, Appl
202	5	2.3	33	2	US-08-467-881A-84	Sequence 84, Appl
203	5	2.3	34	1	US-08-118-270-184	Sequence 184, App
204	5	2.3	34	1	US-08-257-528B-45	Sequence 45, Appl
205	5	2.3	34	2	US-08-460-602A-45	Sequence 45, Appl
206	5	2.3	34	1	US-08-463-966A-45	Sequence 45, Appl
207	5	2.3	34	1	US-08-465-217A-45	Sequence 45, Appl
208	5	2.3	34	2	US-08-464-329A-45	Sequence 45, Appl
209	5	2.3	34	2	US-08-462-507A-45	Sequence 45, Appl
210	5	2.3	34	2	US-08-467-881A-45	Sequence 45, Appl
211	5	2.3	34	4	PCT-US93-08528-184	Sequence 184, App
212	5	2.3	36	1	US-08-781-020-12	Sequence 12, Appl
213	5	2.3	37	2	US-08-212-236-10	Sequence 10, Appl
214	5	2.3	43	2	US-08-533-996A-2	Sequence 2, Appl
215	5	2.3	45	1	US-08-389-360-2	Sequence 102, App
216	5	2.3	45	2	US-08-056-200-102	Sequence 14, Appl
217	5	2.3	45	1	US-08-770-544-14	Sequence 102, App
218	5	2.3	48	2	US-08-800-644-102	Sequence 49, Appl
219	5	2.3	48	3	US-08-460-890A-49	Sequence 49, Appl
220	5	2.3	49	2	US-08-167-641C-49	Sequence 48, Appl
221	5	2.3	49	3	US-08-460-890A-48	Sequence 48, Appl
222	5	2.3	49	3	US-08-167-641C-48	Sequence 6, Appl
223	5	2.3	49	4	PCT-US92-07813-6	Sequence 7, Appl
224	5	2.3	52	1	US-08-381-666-1	Sequence 1, Appl
225	5	2.3	52	2	US-08-309-420-7	Sequence 7, Appl
226	5	2.3	52	1	US-08-309-419-7	Sequence 7, Appl
227	5	2.3	52	4	PCT-US95-11856-7	Sequence 144, App
228	5	2.3	53	4	PCT-US95-11878-7	Sequence 79, Appl
229	5	2.3	53	2	US-07-857-060-1	Sequence 54, Appl
230	5	2.3	54	3	US-08-726-306A-144	Sequence 38, Appl
231	5	2.3	54	1	US-08-630-916A-79	Sequence 38, Appl
232	5	2.3	54	2	US-08-237-401A-38	Sequence 28, Appl
233	5	2.3	54	3	US-08-456-647B-38	Sequence 1, Appl
234	5	2.3	58	1	US-08-691-814B-28	Sequence 144, App
235	5	2.3	62	2	US-08-381-049-1	Sequence 79, Appl
236	5	2.3	63	2	US-08-381-041A-1	Sequence 54, Appl
237	5	2.3	63	1	US-08-257-528B-91	Sequence 38, Appl
238	5	2.3	63	2	US-08-460-602A-91	Sequence 38, Appl
239	5	2.3	63	1	US-08-463-966A-91	Sequence 28, Appl
240	5	2.3	63	1	US-08-465-217A-91	Sequence 1, Appl
241	5	2.3	63	2	US-08-464-329A-91	Sequence 91, Appl
242	5	2.3	63	2	US-08-462-507A-91	Sequence 91, Appl
243	5	2.3	63	2	US-08-467-881A-91	Sequence 91, Appl
244	5	2.3	63	1	US-08-464-329A-91	Sequence 91, Appl
245	5	2.3	63	2	US-08-465-217A-91	Sequence 91, Appl
246	5	2.3	64	2	US-08-462-507A-91	Sequence 92, Appl
247	5	2.3	64	1	US-08-460-602A-92	Sequence 92, Appl

#					Reference	Entry
248	2.3	5	64	1	US-08-463-966A-92	Sequence 92, Appl
249	2.3	5	64	1	US-08-465-217A-92	Sequence 92, Appl
250	2.3	5	64	2	US-08-464-329A-92	Sequence 92, Appl
251	2.3	5	64	2	US-08-462-507A-92	Sequence 92, Appl
252	2.3	5	64	2	US-08-467-881A-92	Sequence 92, Appl
253	2.3	5	66	1	US-08-262-037-50	Sequence 50, Appl
254	2.3	5	68	1	US-08-330-638D-2	Sequence 2, Appl
255	2.3	5	68	1	US-08-606-789-8	Sequence 8, Appli
256	2.3	5	68	2	US-09-111-348-8	Sequence 8, Appl
257	2.3	5	68	2	US-08-511-485-28	Sequence 28, Appl
258	2.3	5	68	2	US-08-906-746A-2	Sequence 2, Appl
259	2.3	5	69	1	US-08-350-884-74	Sequence 74, Appl
260	2.3	5	69	1	US-08-709-173-74	Sequence 74, Appl
261	2.3	5	69	2	US-08-709-177-74	Sequence 74, Appl
262	2.3	5	69	4	PCT-US95-06406A-6	Sequence 6, Appli
263	2.3	5	70	1	US-08-315-695-3	Sequence 3, Appli
264	2.3	5	72	1	US-08-280-443-11	Sequence 11, Appl
265	2.3	5	72	1	US-08-457-459-11	Sequence 11, Appl
266	2.3	5	72	1	US-08-555-678-11	Sequence 11, Appl
267	2.3	5	72	4	PCT-US95-02275-11	Sequence 11, Appll
268	2.3	5	74	1	US-08-381-037-1	Sequence 1, Appli
269	2.3	5	74	4	PCT-US92-00282-24	Sequence 24, Appl
270	2.3	5	75	1	US-08-381-054-1	Sequence 1, Appll
271	2.3	5	76	1	US-07-881-075-6	Sequence 6, Appll
272	2.3	5	76	1	US-08-120-827-6	Sequence 6, Appll
273	2.3	5	76	1	US-08-478-675-6	Sequence 6, Appli
274	2.3	5	81	2	US-08-469-667-14	Sequence 14, Appl
275	2.3	5	81	2	US-09-014-969-32	Sequence 32, Appl
276	2.3	5	81	4	PCT-US95-07289-14	Sequence 14, Appl
277	2.3	5	83	3	US-08-665-259-14	Sequence 14, Appl
278	2.3	5	84	2	US-08-901-306-4	Sequence 4, Appll
279	2.3	5	85	1	US-08-381-163-1	Sequence 1, Appll
280	2.3	5	88	5	5206352-2	Patent No. 5206352
281	2.3	5	90	2	US-08-718-538-7	Sequence 7, Appll
282	2.3	5	91	2	US-08-633-682-5	Sequence 5, Appl
283	2.3	5	91	1	US-08-936-772-5	Sequence 5, Appl
284	2.3	5	93	1	US-08-581-040-1	Sequence 1, Appll
285	2.3	5	93	1	US-08-350-884-76	Sequence 76, Appl
286	2.3	5	93	2	US-08-709-173-76	Sequence 76, Appl
287	2.3	5	93	2	US-08-709-177-76	Sequence 76, Appl
288	2.3	5	95	2	US-09-047-125-30	Sequence 30, Appl
289	2.3	5	99	2	US-08-858-767-28	Sequence 28, Appl
290	2.3	5	99	2	US-08-863-028-28	Sequence 28, Appl
291	2.3	5	99	2	US-08-710-749-9	Sequence 9, Appli
292	2.3	5	99	2	US-08-710-749-10	Sequence 10, Appl
293	2.3	5	99	2	US-08-710-749-11	Sequence 11, Appl
294	2.3	5	99	2	US-08-710-749-12	Sequence 12, Appl
295	2.3	5	101	2	US-08-710-749-1	Sequence 1, Appll
296	2.3	5	101	2	US-08-710-749-2	Sequence 2, Appll
297	2.3	5	101	2	US-08-710-749-4	Sequence 4, Appll
298	2.3	5	102	2	US-08-710-749-7	Sequence 7, Appll
299	2.3	5	104	1	US-08-710-749-8	Sequence 8, Appli
300	2.3	5	105	2	US-08-315-269-2	Sequence 2, Appll
301	2.3	5	109	2	US-08-381-034-1	Sequence 1, Appll
302	2.3	5	109	2	US-08-352-324A-2	Sequence 2, Appll
303	2.3	5	109	2	US-08-352-324A-9	Sequence 9, Appll
304	2.3	5	110	2	US-08-862-607-2	Sequence 2, Appll
305	2.3	5	109	2	US-08-862-607-9	Sequence 9, Appll
306	2.3	5	109	4	PCT-US95-01780-4	Sequence 4, Appli
307	2.3	5	109	4	PCT-US95-16144-9	Sequence 9, Appll
308	2.3	5	110	2	US-08-362-780-26	Sequence 26, Appl
309	2.3	5	111	1	US-08-381-050-1	Sequence 1, Appll
310	2.3	5	112	1	US-08-446-919A-5	Sequence 5, Appll
311	2.3	5	113	1	US-08-381-057-1	Sequence 1, Appli
312	2.3	5	114	2	US-08-352-324A-3	Sequence 3, Appll
313	2.3	5	114	1	US-08-862-607-3	Sequence 3, Appll
314	2.3	5	114	2	US-08-468-819-4	Sequence 4, Appll
315	2.3	5	114	4	PCT-US95-16144-3	Sequence 3, Appll
316	2.3	5	115	4	PCT-US95-16144-9	Sequence 9, Appll
317	2.3	5	115	2	US-08-111-939-22	Sequence 22, Appl
318	2.3	5	116	2	US-08-853-659A-42	Sequence 42, Appl
319	2.3	5	117	2	US-08-557-309B-48	Sequence 48, Appl
320	2.3	5	119	1	US-08-383-639-1	Sequence 1, Appli
321	2.3	5	119	2	US-08-347-563A-25	Sequence 25, Appl
322	2.3	5	120	2	US-08-232-087A-8	Sequence 8, Appl
323	2.3	5	122	1	US-08-350-884-72	Sequence 72, Appl
324	2.3	5	122	1	US-08-266-451B-18	Sequence 18, Appl
325	2.3	5	122	1	US-08-497-025-5	Sequence 5, Appl
326	2.3	5	122	1	US-08-709-173-72	Sequence 72, Appl
327	2.3	5	122	2	US-08-748-725-18	Sequence 18, Appl
328	2.3	5	122	2	US-08-709-177-72	Sequence 72, Appl
329	2.3	5	122	5	5169835-12	Patent No. 5169835
330	2.3	5	125	1	US-08-478-039-69	Sequence 69, Appl
331	2.3	5	125	1	US-08-476-349A-69	Sequence 69, Appl
332	2.3	5	126	1	US-08-383-649-1	Sequence 1, Appl
333	2.3	5	127	2	US-08-853-659A-56	Sequence 56, Appl
334	2.3	5	128	1	US-08-384-183-1	Sequence 1, Appl
335	2.3	5	131	1	US-08-384-492-1	Sequence 1, Appll
336	2.3	5	131	2	US-08-394-189B-21	Sequence 21, Appl
337	2.3	5	132	1	US-08-392-419-4	Sequence 4, Appl
338	2.3	5	133	1	US-08-253-877C-10	Sequence 10, Appl
339	2.3	5	133	1	US-08-452-164A-10	Sequence 10, Appl
340	2.3	5	133	3	US-08-603-024-4	Sequence 4, Appll
341	2.3	5	135	1	US-08-383-650-1	Sequence 1, Appll
342	2.3	5	136	1	US-08-426-627-17	Sequence 17, Appl
343	2.3	5	136	2	US-08-580-545B-8	Sequence 8, Appll
344	2.3	5	138	1	US-08-384-292-1	Sequence 1, Appll
345	2.3	5	138	2	US-08-350-884-78	Sequence 78, Appl
346	2.3	5	138	1	US-08-468-709B-2	Sequence 2, Appll
347	2.3	5	138	2	US-08-468-709B-12	Sequence 12, Appl
348	2.3	5	138	2	US-08-709-173-78	Sequence 78, Appl
349	2.3	5	138	2	US-08-241-664B-2	Sequence 2, Appll
350	2.3	5	138	2	US-08-241-664B-12	Sequence 12, Appl
351	2.3	5	138	2	US-08-709-177-78	Sequence 78, Appl
352	2.3	5	138	4	PCT-US93-03936-2	Sequence 2, Appll
353	2.3	5	139	2	US-08-805-918-2	Sequence 2, Appll
354	2.3	5	140	3	US-09-053-197A-34	Sequence 34, Appl
355	2.3	5	141	1	US-08-438-123-7	Sequence 7, Appl
356	2.3	5	141	2	US-08-383-632-1	Sequence 1, Appll
357	2.3	5	143	5	5169835-18	Patent No. 5169835
358	2.3	5	144	2	US-08-442-916-1	Sequence 1, Appll
359	2.3	5	144	1	US-08-905-694-1	Sequence 1, Appll
360	2.3	5	144	4	PCT-US93-07213-11	Sequence 11, Appl
361	2.3	5	144	5	5169835-8	Patent No. 5169835
362	2.3	5	145	2	US-08-832-535-2	Sequence 2, Appl
363	2.3	5	146	2	US-08-399-021-3	Sequence 3, Appl
364	2.3	5	146	1	US-08-678-369-4	Sequence 4, Appl
365	2.3	5	146	1	US-08-678-369-7	Sequence 7, Appl
366	2.3	5	146	1	US-08-678-369-8	Sequence 8, Appl
367	2.3	5	146	2	US-08-788-943A-1	Sequence 1, Appl
368	2.3	5	146	2	US-08-788-943A-2	Sequence 2, Appl
369	2.3	5	146	2	US-08-788-943A-3	Sequence 3, Appl
370	2.3	5	146	2	US-08-788-943A-4	Sequence 4, Appl
371	2.3	5	146	2	US-08-788-943A-5	Sequence 5, Appl
372	2.3	5	146	2	US-08-788-943A-6	Sequence 6, Appl
373	2.3	5	146	2	US-08-788-943A-7	Sequence 7, Appl
374	2.3	5	146	2	US-08-788-943A-8	Sequence 8, Appl
375	2.3	5	146	2	US-08-788-943A-9	Sequence 9, Appl
376	2.3	5	146	2	US-08-788-943A-10	Sequence 10, Appl
377	2.3	5	146	2	US-08-788-943A-11	Sequence 11, Appl
378	2.3	5	146	2	US-08-788-943A-12	Sequence 12, Appl
379	2.3	5	146	2	US-08-788-943A-13	Sequence 13, Appl
380	2.3	5	146	2	US-08-823-104-3	Sequence 3, Appl
381	2.3	5	146	2	US-08-823-104-4	Sequence 4, Appl
382	2.3	5	146	2	US-08-823-104-5	Sequence 5, Appl
383	2.3	5	146	2	US-08-823-104-6	Sequence 6, Appl
384	2.3	5	146	2	US-08-823-104-7	Sequence 7, Appl
385	2.3	5	146	2	US-08-823-104-8	Sequence 8, Appl
386	2.3	5	146	2	US-08-823-104-9	Sequence 9, Appl
387	2.3	5	146	2	US-08-823-104-10	Sequence 10, Appl
388	2.3	5	146	2	US-08-823-104-11	Sequence 11, Appl
389	2.3	5	146	2	US-08-823-104-12	Sequence 12, Appl
390	2.3	5	146	2	US-08-823-104-13	Sequence 13, Appl
391	2.3	5	146	2	US-08-823-104-14	Sequence 14, Appl
392	2.3	5	146	2	US-08-823-104-15	Sequence 15, Appl
393	2.3	5	146	2	US-08-823-104-16	Sequence 16, Appl

Left column

Index	Tm	Len	Patent	Sequence
394	2.3	5	US-09-047-243-4	Sequence 4, Appl1
395	2.3	5	US-09-047-243-7	Sequence 7, Appl1
396	2.3	5	US-09-047-243-8	Sequence 8, Appl1
397	2.3	5	US-08-804-668-1	Sequence 1, Appl1
398	2.3	5	US-09-003-081-6	Sequence 6, Appl1
399	2.3	5	US-08-648-262-6	Sequence 6, Appl1
400	2.3	5	US-08-648-263-6	Sequence 1, Appl1
401	2.3	5	PCT-US96-00952-1	Sequence 2, Appl1
402	2.3	5	PCT-US96-00952-2	Sequence 3, Appl1
403	2.3	5	PCT-US96-00952-3	Sequence 4, Appl1
404	2.3	5	PCT-US96-00952-4	Sequence 5, Appl1
405	2.3	5	PCT-US96-00952-5	Sequence 6, Appl1
406	2.3	5	PCT-US96-00952-6	Sequence 7, Appl1
407	2.3	5	PCT-US96-00952-7	Sequence 8, Appl1
408	2.3	5	PCT-US96-00952-8	Sequence 9, Appl1
409	2.3	5	PCT-US96-00952-9	Sequence 10, Appl1
410	2.3	5	PCT-US96-00952-10	Sequence 11, Appl1
411	2.3	5	PCT-US96-00952-11	Sequence 3, Appl1
412	2.3	5	US-08-429-362-3	Sequence 17, Appl1
413	2.3	5	US-08-823-104-17	Sequence 18, Appl1
414	2.3	5	US-08-823-104-18	Sequence 2, Appl1
415	2.3	5	US-08-916-043-2	Sequence 9, Appl1
416	2.3	5	5212286-6	Patent No. 5212286
417	2.3	5	5212286-6	Sequence 3, Appl1
418	2.3	5	US-08-756-749C-3	Sequence 15, Appl1
419	2.3	5	US-08-984-172-3	Sequence 3, Appl1
420	2.3	5	US-08-328-322-15	Sequence 6, Appl1
421	2.3	5	US-08-828-832-3	Sequence 16, Appl1
422	2.3	5	US-08-670-186-6	Sequence 3, Appl1
423	2.3	5	US-08-162-402B-16	Sequence 40, Appl1
424	2.3	5	US-08-918-723-3	Sequence 6, Appl1
425	2.3	5	US-09-237-507-3	Sequence 4, Appl1
426	2.3	5	US-08-233-788A-40	Sequence 6, Appl1
427	2.3	5	US-08-918-723-4	Sequence 4, Appl1
428	2.3	5	US-08-347-563A-6	Sequence 5, Appl1
429	2.3	5	US-09-237-507-4	Sequence 5, Appl1
430	2.3	5	US-08-292-345B-6	Sequence 4, Appl1
431	2.3	5	US-08-540-242A-4	Sequence 22, Appl1
432	2.3	5	US-08-347-563A-4	Sequence 1, Appl1
433	2.3	5	US-08-292-345B-4	Sequence 1, Appl1
434	2.3	5	US-08-648-262-5	Sequence 17, Appl1
435	2.3	5	US-08-648-263-5	Sequence 5, Appl1
436	2.3	5	PCT-US96-01471-4	Sequence 1, Appl1
437	2.3	5	US-08-820-170A-22	Sequence 32, Appl1
438	2.3	5	US-08-742-009-1	Sequence 2, Appl1
439	2.3	5	US-08-918-723-1	Sequence 15, Appl1
440	2.3	5	US-09-237-507-1	Sequence 15, Appl1
441	2.3	5	US-09-172-110-1	Sequence 1, Appl1
442	2.3	5	US-08-328-322-17	Patent No. 5463025
443	2.3	5	US-08-809-267-5	Sequence 27, Appl1
444	2.3	5	PCT-US95-13662A-5	Sequence 28, Appl1
445	2.3	5	US-08-449-922-1	Sequence 29, Appl1
446	2.3	5	US-08-049-473-27	Sequence 27, Appl1
447	2.3	5	US-08-049-473-28	Sequence 28, Appl1
448	2.3	5	US-08-049-473-29	Sequence 29, Appl1
449	2.3	5	US-08-044-621D-32	Sequence 4, Appl1
450	2.3	5	US-07-929-580B-2	Sequence 4, Appl1
451	2.3	5	US-08-709-912-15	Sequence 6, Appl1
452	2.3	5	US-09-047-370-15	Sequence 27, Appl1
453	2.3	5	US-08-791-522-1	Sequence 28, Appl1
454	2.3	5	5463025-4	Sequence 29, Appl1
455	2.3	5	US-08-049-473-27	Sequence 5, Appl1
456	2.3	5	US-08-049-473-28	Sequence 22, Appl1
457	2.3	5	US-08-312-648-27	Sequence 3, Appl1
458	2.3	5	US-08-312-648-28	
459	2.3	5	US-08-312-648-29	
460	2.3	5	US-08-418-444A-4	
461	2.3	5	US-08-418-444A-4	
462	2.3	5	US-08-418-444A-6	
463	2.3	5	PCT-US94-04190-27	
464	2.3	5	PCT-US94-04190-28	
465	2.3	5	PCT-US94-04190-29	
466	2.3	5	US-08-558-823-5	

Right column

Index	Tm	Len	Pos	Count	Patent	Sequence
467	2.3	5	189	2	US-09-096-082-3	Sequence 3, Appl1
468	2.3	5	192	1	US-08-086-428B-86	Sequence 86, Appl1
469	2.3	5	192	1	US-08-086-428B-87	Sequence 87, Appl1
470	2.3	5	192	1	US-08-086-428B-88	Sequence 88, Appl1
471	2.3	5	192	2	US-08-086-428B-89	Sequence 89, Appl1
472	2.3	5	192	2	US-08-468-570-86	Sequence 86, Appl1
473	2.3	5	192	2	US-08-468-570-87	Sequence 87, Appl1
474	2.3	5	192	2	US-08-468-570-88	Sequence 88, Appl1
475	2.3	5	192	2	US-08-468-570-89	Sequence 89, Appl1
476	2.3	5	192	2	US-08-290-665A-86	Sequence 86, Appl1
477	2.3	5	192	2	US-08-290-665A-87	Sequence 87, Appl1
478	2.3	5	192	2	US-08-290-665A-88	Sequence 88, Appl1
479	2.3	5	192	2	US-08-290-665A-89	Sequence 89, Appl1
480	2.3	5	192	4	PCT-US95-10398-86	Sequence 86, Appl1
481	2.3	5	192	4	PCT-US95-10398-87	Sequence 87, Appl1
482	2.3	5	192	4	PCT-US95-10398-88	Sequence 88, Appl1
483	2.3	5	192	4	PCT-US95-10398-89	Sequence 89, Appl1
484	2.3	5	197	1	US-08-044-621D-29	Sequence 29, Appl1
485	2.3	5	197	1	US-08-567-505-1	Sequence 1, Appl1
486	2.3	5	197	2	US-08-709-912-9	Sequence 9, Appl1
487	2.3	5	197	2	US-08-779-87D-1	Sequence 1, Appl1
488	2.3	5	197	2	US-09-047-370-9	Sequence 2, Appl1
489	2.3	5	199	1	US-08-017-522A-4	Sequence 4, Appl1
490	2.3	5	199	1	US-08-115-680-2	Sequence 2, Appl1
491	2.3	5	199	1	US-07-941-372-2	Sequence 67, Appl1
492	2.3	5	199	1	US-08-350-884-67	Sequence 4, Appl1
493	2.3	5	199	1	US-07-949-516A-4	Sequence 4, Appl1
494	2.3	5	199	1	US-07-949-173-67	Sequence 67, Appl1
495	2.3	5	199	1	US-08-792-019B-8	Sequence 8, Appl1
496	2.3	5	199	2	US-08-602-262-2	Sequence 2, Appl1
497	2.3	5	199	1	US-08-814-459-4	Sequence 4, Appl1
498	2.3	5	199	2	US-08-709-177-67	Sequence 67, Appl1
499	2.3	5	199	4	PCT-US93-08247-2	Sequence 2, Appl1
500	2.3	5	200	2	US-09-031-485-12	Sequence 12, Appl1
501	2.3	5	200	3	US-08-847-429A-12	Sequence 12, Appl1
502	2.3	5	200	2	US-09-082-737-4	Sequence 4, Appl1
503	2.3	5	202	1	US-08-350-884-1	Sequence 1, Appl1
504	2.3	5	202	1	US-08-350-884-65	Sequence 65, Appl1
505	2.3	5	202	1	US-08-709-173-1	Sequence 1, Appl1
506	2.3	5	202	1	US-08-709-173-65	Sequence 65, Appl1
507	2.3	5	202	2	US-08-531-525-24	Sequence 24, Appl1
508	2.3	5	202	2	US-08-709-177-1	Sequence 1, Appl1
509	2.3	5	202	2	US-08-709-177-65	Sequence 65, Appl1
510	2.3	5	202	1	US-08-718-270A-24	Sequence 24, Appl1
511	2.3	5	203	2	US-08-598-873-4	Sequence 4, Appl1
512	2.3	5	204	2	US-08-446-925-7	Sequence 7, Appl1
513	2.3	5	204	2	US-09-146-331-7	Sequence 7, Appl1
514	2.3	5	204	2	US-08-896-885-7	Sequence 7, Appl1
515	2.3	5	204	2	US-08-576-626A-40	Sequence 40, Appl1
516	2.3	5	208	1	US-08-381-881-4	Sequence 4, Appl1
517	2.3	5	211	2	US-08-631-607-4	Sequence 4, Appl1
518	2.3	5	213	3	US-09-008-303-2	Sequence 2, Appl1
519	2.3	5	215	2	US-08-984-172-1	Sequence 1, Appl1
520	2.3	5	215	2	US-08-997-080-117	Sequence 117, App
521	2.3	5	216	1	US-08-997-362-117	Sequence 117, App
522	2.3	5	216	2	US-08-455-543A-43	Sequence 43, Appl1
523	2.3	5	216	2	US-08-223-305C-43	Sequence 43, Appl1
524	2.3	5	218	1	US-08-149-097D-39	Sequence 39, Appl1
525	2.3	5	218	2	US-07-607-538C-4	Sequence 4, Appl1
526	2.3	5	218	2	US-08-126-593A-4	Sequence 4, Appl1
527	2.3	5	219	2	US-08-454-039A-4	Sequence 4, Appl1
528	2.3	5	219	2	US-08-162-402B-4	Sequence 4, Appl1
529	2.3	5	219	2	US-08-455-543A-44	Sequence 44, Appl1
530	2.3	5	223	1	US-08-223-305C-44	Sequence 44, Appl1
531	2.3	5	224	4	US-08-149-097D-40	Sequence 40, Appl1
532	2.3	5	227	1	US-08-896-410-4	Sequence 4, Appl1
533	2.3	5	228	1	PCT-US91-09055-4	Sequence 5, Appl1
534	2.3	5	228	1	US-08-287-959-5	Sequence 4, Appl1
535	2.3	5	228	1	US-08-414-625-4	Sequence 7, Appl1
536	2.3	5			US-08-278-091-7	Sequence 7, Appl1
537	2.3	5			US-08-483-859-7	
538	2.3	5			US-08-472-173-7	
539	2.3	5				

Left column

#					Application / Patent No.	Description
540	5	2.3	228	2	US-08-487-167-7	Sequence 7, Appli
541	5	2.3	228	2	US-08-482-816-7	Sequence 7, Appli
542	5	2.3	228	2	US-08-296-149-7	Sequence 7, Appli
543	5	2.3	228	2	US-08-801-499-7	Sequence 7, Appli
544	5	2.3	228	3	US-08-615-271-7	Sequence 7, Appli
545	5	2.3	228	2	US-09-074-660-7	Sequence 7, Appli
546	5	2.3	228	3	US-09-074-659-7	Sequence 7, Appli
547	5	2.3	229	1	US-08-028-463-2	Sequence 2, Appli
548	5	2.3	229	1	US-08-383-750-6	Sequence 6, Appli
549	5	2.3	229	2	US-08-461-836-2	Sequence 2, Appli
550	5	2.3	229	2	US-08-684-687-2	Sequence 2, Appli
551	5	2.3	229	2	US-08-121-436A-4	Sequence 4, Appli
552	5	2.3	229	3	US-08-801-740-3	Sequence 3, Appli
553	5	2.3	229	2	US-08-352-678-6	Sequence 6, Appli
554	5	2.3	230	1	US-08-052-205-11	Sequence 11, Appl
555	5	2.3	230	2	US-08-595-974-11	Sequence 11, Appl
556	5	2.3	230	2	US-08-637-759B-457	Sequence 457, App
557	5	2.3	230	2	US-08-871-355A-457	Sequence 457, App
558	5	2.3	230	4	PCT-US93-09636-6	Sequence 6, Appli
559	5	2.3	231	2	US-08-902-655A-6	Sequence 6, Appli
560	5	2.3	232	2	US-08-446-345-38	Sequence 38, Appl
561	5	2.3	233	1	US-08-800-751-2	Sequence 2, Appli
562	5	2.3	233	2	US-08-738-413B-9	Sequence 9, Appli
563	5	2.3	248	5	5169835-15	Patent No. 5169835
564	5	2.3	248	5	5182262-8	Patent No. 5182262
565	5	2.3	249	3	US-08-486-099-102	Sequence 102, App
566	5	2.3	250	2	US-08-484-223B-102	Sequence 102, App
567	5	2.3	251	2	US-08-360-107A-112	Sequence 112, App
568	5	2.3	251	2	US-08-469-537A-80	Sequence 80, Appl
569	5	2.3	252	4	US-08-469-537A-81	Sequence 81, Appl
570	5	2.3	252	2	PCT-US96-01314-59	Sequence 59, Appl
571	5	2.3	252	1	US-08-052-205-9	Sequence 9, Appli
572	5	2.3	252	2	US-08-595-974-9	Sequence 9, Appli
573	5	2.3	252	1	US-08-299-567-5	Sequence 5, Appli
574	5	2.3	255	1	US-08-462-949-3	Sequence 3, Appli
575	5	2.3	255	2	US-08-023-764B-3	Sequence 11, Appl
576	5	2.3	255	2	US-08-103-170-11	Sequence 2, Appli
577	5	2.3	255	4	US-08-240-124-2	Sequence 2, Appli
578	5	2.3	252	1	PCT-US96-01314-59	Sequence 2, Appli
579	5	2.3	257	2	US-08-453-943-2	Sequence 9, Appli
580	5	2.3	258	1	US-09-057-121-2	Sequence 9, Appli
581	5	2.3	258	2	US-08-460-512-4	Sequence 4, Appli
582	5	2.3	258	2	US-08-853-659A-43	Sequence 43, Appl
583	5	2.3	259	2	US-08-685-992-16	Sequence 16, Appl
584	5	2.3	260	2	US-08-144-925-16	Sequence 16, Appl
585	5	2.3	262	4	US-08-310-912A-106	Sequence 106, App
586	5	2.3	262	2	PCT-US95-04570-106	Sequence 106, App
587	5	2.3	262	4	PCT-US95-04589-106	Sequence 106, App
588	5	2.3	262	2	US-08-467-265-16	Sequence 16, Appl
589	5	2.3	262	2	US-08-744-026-3	Sequence 3, Appli
590	5	2.3	258	1	US-09-102-732-3	Sequence 7, Appli
591	5	2.3	259	3	US-08-402-804-8	Sequence 8, Appli
592	5	2.3	260	1	US-09-172-977-1	Sequence 1, Appli
593	5	2.3	262	2	US-08-738-944-50	Sequence 50, Appl
594	5	2.3	262	1	US-08-744-026-4	Sequence 1, Appli
595	5	2.3	262	2	US-08-790-137-3	Sequence 3, Appli
596	5	2.3	262	2	US-08-790-137-4	Sequence 4, Appli
597	5	2.3	262	2	US-08-681-151-4	Sequence 5, Appli
598	5	2.3	262	2	US-09-102-732-4	Sequence 4, Appli
599	5	2.3	262	3	US-07-857-224B-62	Sequence 62, Appli
600	5	2.3	262	2	US-08-824-874-4	Sequence 4, Appli
601	5	2.3	262	3	US-08-807-151-4	Sequence 4, Appli
602	5	2.3	263	2	US-08-790-137-4	Sequence 5, Appli
603	5	2.3	263	3	US-08-824-874-5	Sequence 5, Appli
604	5	2.3	263	3	US-09-035-706-5	Sequence 5, Appli
605	5	2.3	263	3	US-08-955-841-5	Sequence 5, Appli
606	5	2.3	266	3	US-08-807-151-5	Sequence 4, Appli
607	5	2.3	266	5	5175384-11	Patent No. 5175384
608	5	2.3	267	5	US-07-857-224B-37	Sequence 37, Appl
609	5	2.3	267	2	US-08-793-490-11	Sequence 11, Appl
610	5	2.3	268	3	US-08-836-134-22	Sequence 22, Appli
611	5	2.3	269	1	US-08-241-766-6	Sequence 6, Appli
612	5	2.3	269	1	US-08-241-766-7	Sequence 7, Appli

Right column

#					Application / Patent No.	Description
613	5	2.3	271	3	US-08-536-891A-1	Sequence 1, Appl1
614	5	2.3	272	1	US-08-350-884-84	Sequence 84, Appl
615	5	2.3	272	2	US-08-709-173-84	Sequence 84, Appl
616	5	2.3	272	2	US-08-852-743-6	Sequence 6, Appl1
617	5	2.3	272	2	US-08-160-524A-7	Sequence 7, Appl1
618	5	2.3	272	1	US-08-709-177-84	Sequence 84, Appl
619	5	2.3	273	2	US-08-215-928A-2	Sequence 2, Appl1
620	5	2.3	273	2	US-08-446-920-2	Sequence 2, Appl1
621	5	2.3	275	2	US-08-701-191A-38	Sequence 38, Appl
622	5	2.3	276	2	US-08-070-301-11	Sequence 11, Appl
623	5	2.3	277	2	US-08-469-537A-57	Sequence 57, Appl
624	5	2.3	277	1	US-07-941-414-1	Sequence 1, Appl1
625	5	2.3	278	1	US-08-249-377A-1	Sequence 1, Appl1
626	5	2.3	278	1	US-08-585-595-1	Sequence 1, Appl1
627	5	2.3	278	2	US-08-416-603-6	Sequence 6, Appl1
628	5	2.3	278	1	US-08-749-337-1	Sequence 6, Appl1
629	5	2.3	278	2	US-08-865-203-6	Sequence 6, Appl1
630	5	2.3	278	2	US-07-849-420-6	Sequence 6, Appl1
631	5	2.3	278	3	US-08-522-613-4	Sequence 2, Appl1
632	5	2.3	278	4	PCT-US94-03744-2	Sequence 4, Appl1
633	5	2.3	279	2	US-09-135-658-4	Sequence 8, Appl1
634	5	2.3	279	4	PCT-US91-00899-8	Patent No. 5472855
635	5	2.3	279	5	5472855-6	Sequence 6, Appl1
636	5	2.3	280	1	US-08-312-387B-6	Sequence 6, Appl1
637	5	2.3	280	1	US-08-683-426-6	Sequence 6, Appl1
638	5	2.3	280	1	US-08-683-458-6	Sequence 6, Appl1
639	5	2.3	280	2	US-08-878-360-6	Sequence 3, Appl1
640	5	2.3	284	2	US-08-767-096-3	Sequence 15, Appl
641	5	2.3	285	2	US-08-631-200-15	Sequence 15, Appl
642	5	2.3	285	2	US-08-829-553-15	Sequence 15, Appl
643	5	2.3	285	2	US-08-922-267A-15	Sequence 15, Appl
644	5	2.3	285	2	US-08-936-707A-15	Sequence 15, Appl
645	5	2.3	285	3	US-08-936-706A-15	Sequence 15, Appl
646	5	2.3	285	3	US-09-248-203-15	Sequence 7, Appl1
647	5	2.3	287	2	US-08-622-353-7	Sequence 9, Appl1
648	5	2.3	287	2	US-08-622-352A-9	Sequence 34, Appl
649	5	2.3	290	2	US-08-576-626A-34	Sequence 6, Appl1
650	5	2.3	292	1	US-08-286-888B-6	Sequence 5, Appl1
651	5	2.3	293	2	US-08-446-925-5	Sequence 5, Appl1
652	5	2.3	293	2	US-09-146-331-5	Sequence 5, Appl1
653	5	2.3	293	2	US-08-896-885-5	Sequence 31, Appl
654	5	2.3	293	4	PCT-US96-10521-31	Sequence 7, Appl1
655	5	2.3	294	3	US-09-057-969-7	Sequence 14, Appl
656	5	2.3	294	1	US-07-745-382-14	Sequence 14, Appl
657	5	2.3	296	1	US-07-921-848-14	Sequence 4, Appl1
658	5	2.3	296	1	US-08-115-680-4	Sequence 4, Appl1
659	5	2.3	296	1	US-07-941-372-4	Sequence 14, Appl
660	5	2.3	296	1	US-08-165-301A-14	Sequence 6, Appl1
661	5	2.3	296	1	US-08-507-431-6	Sequence 4, Appl1
662	5	2.3	296	2	US-08-836-854-4	Sequence 4, Appl1
663	5	2.3	296	4	PCT-US93-08247-4	Sequence 14, Appl
664	5	2.3	296	4	PCT-US94-14179-14	Sequence 6, Appl1
665	5	2.3	297	1	US-07-866-560-6	Sequence 6, Appl1
666	5	2.3	297	2	US-08-077-673-6	Sequence 6, Appl1
667	5	2.3	299	1	US-08-478-992-6	Sequence 66, Appl
668	5	2.3	299	2	US-08-350-884-66	Sequence 68, Appl
669	5	2.3	299	2	US-08-350-884-68	Sequence 66, Appl
670	5	2.3	299	2	US-08-709-173-66	Sequence 68, Appl
671	5	2.3	299	1	US-08-709-173-68	Sequence 66, Appl
672	5	2.3	299	2	US-08-709-177-66	Sequence 68, Appl
673	5	2.3	299	2	US-08-709-177-68	Sequence 2, Appl1
674	5	2.3	303	2	US-07-917-111-2	Sequence 3, Appl1
675	5	2.3	303	3	US-07-917-111-3	Sequence 2, Appl1
676	5	2.3	303	1	US-08-479-638-2	Sequence 3, Appl1
677	5	2.3	303	1	US-08-479-638-3	Sequence 2, Appl1
678	5	2.3	303	2	US-08-294-871A-2	Sequence 6, Appl1
679	5	2.3	303	2	US-08-294-871A-6	Sequence 4, Appl1
680	5	2.3	303	2	US-08-294-871A-8	Sequence 8, Appl1
681	5	2.3	303	2	US-08-294-871A-10	Sequence 10, Appl
682	5	2.3	303	2	US-08-294-871A-12	Sequence 12, Appl
683	5	2.3	303	2	US-08-294-871A-14	Sequence 14, Appl
684	5	2.3	303	2	US-08-294-871A-16	Sequence 16, Appl
685	5	2.3	303	2	US-08-294-871A-16	Sequence 16, Appl

No.					Accession	Description
686	2.3	5	303	2	US-08-294-871A-18	Sequence 18, Appl
687	2.3	5	303	2	US-08-294-871A-20	Sequence 20, Appl
688	2.3	5	303	2	US-08-294-871A-22	Sequence 22, Appl
689	2.3	5	303	2	US-08-294-871A-24	Sequence 24, Appl
690	2.3	5	303	2	US-08-294-871A-26	Sequence 26, Appl
691	2.3	5	303	2	US-08-294-871A-28	Sequence 28, Appl
692	2.3	5	303	2	US-08-294-871A-30	Sequence 30, Appl
693	2.3	5	303	2	US-08-294-871A-32	Sequence 32, Appl
694	2.3	5	303	2	US-08-294-871A-34	Sequence 34, Appl
695	2.3	5	303	2	US-08-294-871A-36	Sequence 36, Appl
696	2.3	5	303	2	US-08-294-871A-38	Sequence 38, Appl
697	2.3	5	303	2	US-08-294-871A-40	Sequence 40, Appl
698	2.3	5	303	2	US-08-294-871A-42	Sequence 42, Appl
699	2.3	5	303	2	US-08-294-871A-44	Sequence 44, Appl
700	2.3	5	303	2	US-08-294-871A-46	Sequence 46, Appl
701	2.3	5	303	2	US-08-294-871A-48	Sequence 48, Appl
702	2.3	5	303	2	US-08-294-871A-50	Sequence 50, Appl
703	2.3	5	303	2	US-08-294-871A-52	Sequence 52, Appl
704	2.3	5	303	2	US-08-294-871A-54	Sequence 54, Appl
705	2.3	5	303	2	US-08-294-871A-56	Sequence 56, Appl
706	2.3	5	303	2	US-08-294-871A-58	Sequence 58, Appl
707	2.3	5	303	2	US-08-294-871A-60	Sequence 60, Appl
708	2.3	5	303	2	US-08-294-871A-62	Sequence 62, Appl
709	2.3	5	303	2	US-08-294-871A-64	Sequence 64, Appl
710	2.3	5	303	2	US-08-294-871A-66	Sequence 66, Appl
711	2.3	5	303	2	US-08-294-871A-68	Sequence 68, Appl
712	2.3	5	303	2	US-08-203-532F-2	Sequence 2, Appli
713	2.3	5	303	4	PCT-US95-01882A-2	Sequence 2, Appli
714	2.3	5	304	1	US-08-900-711-5	Sequence 5, Appli
715	2.3	5	304	2	US-08-815-356-2	Sequence 2, Appli
716	2.3	5	304	2	US-08-415-343B-5	Sequence 5, Appli
717	2.3	5	304	2	US-08-762-433-1	Sequence 1, Appli
718	2.3	5	307	2	US-09-001-219-1	Sequence 1, Appli
719	2.3	5	308	1	US-08-457-245-8	Sequence 8, Appli
720	2.3	5	309	2	US-08-405-175A-56	Sequence 56, Appli
721	2.3	5	310	1	US-08-405-175A-7	Sequence 7, Appli
722	2.3	5	310	2	US-08-078-683A-3	Sequence 3, Appli
723	2.3	5	310	1	US-08-433-783-42	Sequence 42, Appli
724	2.3	5	310	2	US-08-484-956-88	Sequence 88, Appli
725	2.3	5	310	2	US-08-757-653-88	Sequence 88, Appli
726	2.3	5	310	2	US-08-488-199-4	Sequence 4, Appli
727	2.3	5	310	2	US-08-337-358-42	Sequence 42, Appli
728	2.3	5	310	4	PCT-US95-07537A-42	Sequence 42, Appli
729	2.3	5	312	2	US-08-414-625-2	Sequence 2, Appli
730	2.3	5	314	1	5434340-7	Patent No. 5434340
731	2.3	5	315	2	US-08-484-956-91	Sequence 91, Appl
732	2.3	5	315	2	US-08-757-653-91	Sequence 91, Appl
733	2.3	5	317	2	US-08-864-799-4	Sequence 4, Appli
734	2.3	5	317	2	US-08-977-847-3	Sequence 5, Appli
735	2.3	5	317	2	US-09-195-021-3	Sequence 3, Appli
736	2.3	5	320	2	US-08-757-653-163	Sequence 163, App
737	2.3	5	320	2	US-08-823-516-61	Sequence 61, Appl
738	2.3	5	321	1	US-08-447-185-1	Sequence 1, Appli
739	2.3	5	322	2	US-08-484-956-89	Sequence 89, Appl
740	2.3	5	322	2	US-08-757-653-89	Sequence 89, Appl
741	2.3	5	322	2	US-08-576-626A-33	Sequence 33, Appl
742	2.3	5	323	2	US-08-540-804-18	Sequence 18, Appl
743	2.3	5	323	2	US-08-218-265-18	Sequence 18, Appl
744	2.3	5	323	2	US-08-521-872-18	Sequence 18, Appl
745	2.3	5	327	2	US-08-896-410-38	Sequence 38, Appl
746	2.3	5	327	1	US-08-926-724-1	Sequence 1, Appli
747	2.3	5	327	2	US-08-148-215A-2	Sequence 2, Appli
748	2.3	5	328	1	US-08-828-242-1	Sequence 1, Appli
749	2.3	5	328	2	US-08-910-927B-1	Sequence 1, Appli
750	2.3	5	329	2	US-08-606-143-1	Sequence 1, Appli
751	2.3	5	329	2	US-08-606-143-3	Sequence 3, Appli
752	2.3	5	331	2	US-08-997-080-182	Sequence 182, App
753	2.3	5	331	2	US-08-997-362-182	Sequence 182, App
759	2.3	5	332	2	US-08-637-763B-8	Sequence 8, Appl1
760	2.3	5	333	2	US-08-853-659A-48	Sequence 48, Appl1
761	2.3	5	334	3	US-08-559-524A-2	Sequence 2, Appl1
762	2.3	5	334	3	US-08-472-595-9	Sequence 9, Appl1
763	2.3	5	334	3	US-08-207-575A-9	Sequence 9, Appl1
764	2.3	5	336	2	US-08-997-080-156	Sequence 156, App
765	2.3	5	336	2	US-08-997-362-156	Sequence 156, Appl
766	2.3	5	337	1	US-08-118-270-33	Sequence 33, Appl
767	2.3	5	337	2	US-08-467-559B-2	Sequence 2, Appl1
768	2.3	5	337	4	PCT-US93-08528-33	Sequence 33, Appl
769	2.3	5	338	2	US-08-502-359A-40	Sequence 40, Appl
770	2.3	5	339	1	US-08-396-957A-4	Sequence 4, Appl1
771	2.3	5	344	1	US-08-400-422-4	Sequence 4, Appl1
772	2.3	5	344	1	US-08-843-993-3	Sequence 3, Appl1
773	2.3	5	344	1	US-08-657-192-3	Sequence 3, Appl1
774	2.3	5	344	2	US-08-689-974-1	Sequence 1, Appl1
775	2.3	5	344	1	US-08-625-377-4	Sequence 4, Appl1
776	2.3	5	344	3	US-09-059-520A-3	Sequence 3, Appl1
777	2.3	5	344	3	US-09-334-275-3	Sequence 3, Appl1
778	2.3	5	344	3	US-08-523-373-5	Sequence 5, Appl1
779	2.3	5	344	5	5210183-2	Patent No. 5210183
780	2.3	5	345	1	US-08-171-382-6	Sequence 6, Appl1
781	2.3	5	345	1	US-08-118-270-43	Sequence 43, Appl1
782	2.3	5	345	1	US-08-309-420-6	Sequence 6, Appl1
783	2.3	5	345	1	US-08-309-419-6	Sequence 6, Appl1
784	2.3	5	345	1	US-08-843-993-1	Sequence 1, Appl1
785	2.3	5	345	3	US-08-282-197C-50	Sequence 50, Appl1
786	2.3	5	345	3	US-09-059-520A-1	Sequence 1, Appl1
787	2.3	5	345	3	US-09-334-275-1	Sequence 1, Appl1
788	2.3	5	345	4	PCT-US93-08528-43	Sequence 43, Appl1
789	2.3	5	345	4	PCT-US95-11856-6	Sequence 6, Appl1
790	2.3	5	345	4	PCT-US95-11878-6	Sequence 6, Appl1
791	2.3	5	346	1	US-08-774-169-3	Sequence 3, Appl1
792	2.3	5	347	1	US-08-052-205-7	Sequence 7, Appl1
793	2.3	5	347	3	US-08-773-870-3	Sequence 3, Appl1
794	2.3	5	347	1	US-08-446-920-10	Sequence 10, Appl1
795	2.3	5	347	1	US-08-595-974-7	Sequence 7, Appl1
796	2.3	5	347	3	US-09-059-369-2	Sequence 2, Appl1
797	2.3	5	348	3	US-08-875-540-13	Sequence 13, Appl1
798	2.3	5	349	1	US-08-872-784-4	Sequence 4, Appl1
799	2.3	5	349	2	US-08-558-823-4	Sequence 4, Appl1
800	2.3	5	349	2	US-08-100-851-4	Sequence 4, Appl1
801	2.3	5	351	2	US-08-159-969-2	Sequence 2, Appl1
802	2.3	5	352	1	US-08-552-142A-4	Sequence 4, Appl1
803	2.3	5	352	2	US-08-910-973-4	Sequence 4, Appl1
804	2.3	5	352	2	US-08-751-174-2	Sequence 2, Appl1
805	2.3	5	352	2	US-08-726-306A-17	Sequence 17, Appl1
806	2.3	5	354	4	PCT-US95-05741-4	Sequence 4, Appl1
807	2.3	5	354	2	US-08-700-013B-9	Sequence 9, Appl1
808	2.3	5	354	2	US-08-394-189B-20	Sequence 20, Appl1
809	2.3	5	355	2	US-08-461-244-2	Sequence 2, Appl1
810	2.3	5	355	3	US-08-666-367B-6	Sequence 6, Appl1
811	2.3	5	356	1	US-07-959-941-2	Sequence 2, Appl1
812	2.3	5	358	2	US-08-259-924-2	Sequence 2, Appl1
813	2.3	5	358	2	US-08-748-485-6	Sequence 6, Appl1
814	2.3	5	359	2	US-07-996-772A-10	Sequence 10, Appl1
815	2.3	5	359	2	US-08-467-568-13	Sequence 13, Appl1
816	2.3	5	359	2	US-08-748-485-3	Sequence 3, Appl1
817	2.3	5	359	2	US-08-748-485-4	Sequence 4, Appl1
818	2.3	5	359	2	US-08-748-485-5	Sequence 5, Appl1
819	2.3	5	359	2	US-08-103-170-4	Sequence 4, Appl1
820	2.3	5	359	2	US-08-103-170-6	Sequence 6, Appl1
821	2.3	5	359	2	US-08-103-170-7	Sequence 7, Appl1
822	2.3	5	359	2	US-09-030-582-13	Sequence 13, Appl1
823	2.3	5	359	2	US-08-875-540-15	Sequence 15, Appl1
824	2.3	5	360	3	US-08-118-270-25	Sequence 25, Appl1
825	2.3	5	360	2	PCT-US93-08528-25	Sequence 25, Appl
826	2.3	5	360	4	US-08-415-751-3	Sequence 3, Appl1
827	2.3	5	361	2	US-08-466-103A-16	Sequence 16, Appl1
828	2.3	5	362	2	US-08-530-950-6	Sequence 6, Appl1
829	2.3	5	363	1	US-08-318-831-6	Sequence 6, Appl1
830	2.3	5	363	2	US-08-650-275-4	Sequence 4, Appl1
831	2.3	5	364	1	US-08-650-275-4	Sequence 4, Appl1

Left column block:

No.	Ver	C	Reference	N	Entry
832	2.3	5	US-09-181-318-4	3	Sequence 4, Appl1
833	2.3	5	US-08-855-518-5	2	Sequence 5, Appl1
834	2.3	5	US-08-169-948B-12	2	Sequence 12, Appl1
835	2.3	5	US-08-448-873-12	2	Sequence 12, Appl1
836	2.3	5	US-08-979-424-3	2	Sequence 3, Appl1
837	2.3	5	US-08-774-169-1	1	Sequence 1, Appl1
838	2.3	5	US-08-855-518-4	2	Sequence 4, Appl1
839	2.3	5	US-08-896-320-1	2	Sequence 1, Appl1
840	2.3	5	US-08-896-320-3	3	Sequence 3, Appl1
841	2.3	5	US-09-032-372-4	2	Sequence 4, Appl1
842	2.3	5	US-08-052-205-4	1	Sequence 4, Appl1
843	2.3	5	US-08-595-974-4	2	Sequence 12, Appl1
844	2.3	5	US-08-378-617A-12	1	Sequence 3, Appl1
845	2.3	5	US-07-813-584A-3	1	Sequence 3, Appl1
846	2.3	5	US-08-330-515-3	2	Sequence 79, Appl1
847	2.3	5	US-08-846-762-79	2	Sequence 24, Appl1
848	2.3	5	US-08-118-270-24	1	Sequence 24, Appl1
849	2.3	5	PCT-US93-08528-24	4	Sequence 14, Appl1
850	2.3	5	US-08-095-726-14	1	Sequence 14, Appl1
851	2.3	5	US-08-096-623A-14	1	Sequence 6, Appl1
852	2.3	5	US-08-737-825-6	1	Patent No. 5180810
853	2.3	5	5180810-1	5	Sequence 3, Appl1
854	2.3	5	US-07-772-087-3	1	Sequence 6, Appl1
855	2.3	5	US-09-057-969-6	3	Sequence 5, Appl1
856	2.3	5	US-08-420-235B-5	2	Sequence 1, Appl1
857	2.3	5	PCT-US95-10194-5	4	Sequence 3, Appl1
858	2.3	5	US-08-858-052-1	2	Sequence 31, Appl1
859	2.3	5	US-08-858-052-3	1	Sequence 58, Appl1
860	2.3	5	US-08-464-523B-31	1	Sequence 58, Appl1
861	2.3	5	US-08-450-257-58	2	Sequence 58, Appl1
862	2.3	5	US-08-450-246-58	1	Sequence 58, Appl1
863	2.3	5	US-08-450-098-58	1	Sequence 4, Appl1
864	2.3	5	US-08-451-233-58	1	Sequence 13, Appl1
865	2.3	5	US-08-450-236-58	1	Sequence 2, Appl1
866	2.3	5	US-08-261-110A-4	2	Sequence 18, Appl1
867	2.3	5	US-08-820-170A-13	2	Sequence 1, Appl1
868	2.3	5	US-07-669-171-2	1	Sequence 1, Appl1
869	2.3	5	US-07-937-609-18	1	Sequence 6, Appl1
870	2.3	5	US-08-132-405-1	1	Patent No. 5168051
871	2.3	5	US-08-395-939A-1	1	Sequence 10, Appl1
872	2.3	5	PCT-US91-01861-1	4	Sequence 23, Appl1
873	2.3	5	PCT-US92-02091-6	4	Sequence 2, Appl1
874	2.3	5	5168051-2	5	Sequence 6, Appl1
875	2.3	5	US-08-244-951A-10	2	Sequence 8, Appl1
876	2.3	5	US-08-389-011-23	3	Sequence 4, Appl1
877	2.3	5	US-08-403-917A-23	3	Sequence 4, Appl1
878	2.3	5	US-08-423-441-2	1	Sequence 4, Appl1
879	2.3	5	US-08-523-373-6	3	Sequence 4, Appl1
880	2.3	5	US-08-530-950-8	2	Sequence 23, Appl1
881	2.3	5	US-08-530-379-4	1	Sequence 9, Appl1
882	2.3	5	US-07-914-281-4	1	Sequence 1, Appl1
883	2.3	5	US-08-990-379-5	1	Sequence 1, Appl1
884	2.3	5	US-08-525-058A-4	2	Sequence 10, Appl1
885	2.3	5	US-08-696-731-4	1	Sequence 92, Appl1
886	2.3	5	US-08-530-290-23	3	Sequence 2, Appl1
887	2.3	5	PCT-US91-00899-9	4	Sequence 2, Appl1
888	2.3	5	PCT-US94-03705-5	4	Sequence 1, Appl1
889	2.3	5	US-08-357-264-1	1	Sequence 2, Appl1
890	2.3	5	US-08-672-514-1	1	Sequence 2, Appl1
891	2.3	5	US-08-990-379-5	1	Sequence 2, Appl1
892	2.3	5	US-08-530-950-10	1	Sequence 2, Appl1
893	2.3	5	US-08-874-186-92	2	Sequence 61, Appl1
894	2.3	5	US-08-974-022-2	1	Sequence 6, Appl1
895	2.3	5	US-08-068-729-4	2	Sequence 2, Appl1
896	2.3	5	US-08-947-726A-2	2	Sequence 2, Appl1
897	2.3	5	US-07-829-954-2	2	Sequence 5, Appl1
898	2.3	5	US-07-994-123-2	1	
899	2.3	5	US-08-421-891-2	2	
900	2.3	5	US-08-406-070-2	2	
901	2.3	5	US-08-282-197C-61	1	
902	2.3	5	US-08-700-013B-6	2	
903	2.3	5	US-08-487-823B-5	1	
904	2.3	5	US-08-997-040-5	2	

Right column block:

No.	Ver	C	Reference	N	Entry
905	2.3	5	US-09-203-237-5	2	Sequence 5, Appl1
906	2.3	5	521296-6	5	Patent No. 521296
907	2.3	5	US-08-381-433A-4	2	Sequence 4, Appl1
908	2.3	5	US-08-381-433A-8	1	Sequence 8, Appl1
909	2.3	5	US-08-399-561-2	1	Sequence 2, Appl1
910	2.3	5	US-08-750-524-1	1	Sequence 1, Appl1
911	2.3	5	US-08-110-286A-2	2	Sequence 2, Appl1
912	2.3	5	US-08-110-286A-6	3	Sequence 6, Appl1
913	2.3	5	PCT-US95-15463-17	4	Sequence 17, Appl1
914	2.3	5	PCT-US95-15923-17	4	Sequence 17, Appl1
915	2.3	5	US-08-997-080-125	2	Sequence 125, App
916	2.3	5	US-08-997-362-125	5	Sequence 125, App
917	2.3	5	5169835-2	5	Patent No. 5169835
918	2.3	5	US-08-487-823B-4	1	Sequence 4, Appl1
919	2.3	5	US-08-997-040-4	2	Sequence 4, Appl1
920	2.3	5	US-09-203-237-4	2	Sequence 4, Appl1
921	2.3	5	US-08-290-731C-10	1	Sequence 10, Appl1
922	2.3	5	5169835-6	5	Patent No. 5169835
923	2.3	5	US-08-732-028-2	2	Sequence 2, Appl1
924	2.3	5	US-08-570-157-5	1	Sequence 5, Appl1
925	2.3	5	5179023-4	1	Patent No. 5179023
926	2.3	5	US-07-602-824A-2	1	Sequence 3, Appl1
927	2.3	5	US-07-602-608-3	1	Sequence 3, Appl1
928	2.3	5	US-07-983-451-2	1	Sequence 3, Appl1
929	2.3	5	US-08-261-578-3	1	Sequence 7, Appl1
929	2.3	5	US-08-261-577-7	1	Sequence 9, Appl1
930	2.3	5	US-08-261-577-9	1	Sequence 23, Appl1
931	2.3	5	US-07-937-609-23	1	Sequence 3, Appl1
932	2.3	5	US-08-919-624-3	2	Sequence 37, Appl1
933	2.3	5	US-08-190-802A-37	1	Sequence 58, Appl1
934	2.3	5	US-08-181-433A-2	1	Sequence 5, Appl1
935	2.3	5	US-08-665-647-5	1	Sequence 9, Appl1
936	2.3	5	US-08-810-116-9	2	Sequence 2, Appl1
937	2.3	5	US-07-930-548A-9	1	Sequence 87, Appl1
938	2.3	5	US-08-691-563C-87	3	Sequence 5, Appl1
939	2.3	5	US-08-374-155A-5	1	Sequence 2, Appl1
940	2.3	5	US-08-785-396-5	1	Sequence 6, Appl1
941	2.3	5	US-08-146-010A-2	1	Sequence 2, Appl1
942	2.3	5	US-08-674-168-2	2	Sequence 4, Appl1
943	2.3	5	US-08-674-168-6	1	Sequence 24, Appl1
944	2.3	5	US-08-261-110A-2	2	Sequence 15, Appl1
945	2.3	5	PCT-US92-09325-4	3	Sequence 2, Appl1
946	2.3	5	PCT-US92-09325-24	1	Sequence 26, Appl1
947	2.3	5	US-08-307-499-15	1	Sequence 28, Appl1
948	2.3	5	US-08-208-108-2	1	Sequence 6, Appl1
949	2.3	5	US-08-993-359-26	3	Sequence 4, Appl1
950	2.3	5	US-08-993-359-28	3	Sequence 30, Appl1
951	2.3	5	US-08-570-157-6	1	Sequence 14, Appl1
952	2.3	5	US-08-795-475-6	1	Sequence 2, Appl1
953	2.3	5	US-08-993-359-30	3	Sequence 67, Appl1
954	2.3	5	US-07-937-609-14	1	Sequence 67, Appl1
955	2.3	5	US-07-937-609-4	1	Sequence 2, Appl1
956	2.3	5	US-07-700-359-4	1	Sequence 22, Appl1
957	2.3	5	PCT-US92-09325-2	4	Sequence 21, Appl1
958	2.3	5	US-07-626-618A-21	1	Sequence 22, Appl1
959	2.3	5	US-07-626-618A-22	1	Sequence 6, Appl1
960	2.3	5	US-08-993-359-28	3	Sequence 4, Appl1
961	2.3	5	US-08-333-977-21	1	Sequence 4, Appl1
962	2.3	5	US-08-333-977-22	2	Sequence 4, Appl1
963	2.3	5	US-08-570-157-6	2	Sequence 2, Appl1
964	2.3	5	US-08-476-063-67	2	Sequence 67, Appl1
965	2.3	5	US-08-306-063-67	2	Sequence 67, Appl1
966	2.3	5	US-08-833-610-2	1	Sequence 2, Appl1
967	2.3	5	US-08-833-485-67	3	Sequence 22, Appl1
968	2.3	5	US-08-508-761B-6	1	Sequence 21, Appl1
969	2.3	5	US-07-937-609-24	1	Sequence 22, Appl1
970	2.3	5	US-08-435-454-4	1	Sequence 6, Appl1
971	2.3	5	US-08-652-972A-4	2	Sequence 4, Appl1
972	2.3	5	US-08-919-145-6	1	Sequence 4, Appl1
973	2.3	5	PCT-US96-06231A-4	4	Sequence 4, Appl1
974	2.3	5	US-08-317-880-2	2	Sequence 2, Appl1
975	2.3	5	US-08-782-396-2	2	Sequence 22, Appl1
976	2.3	5	US-08-993-359-22	3	Sequence 4, Appl1
977	2.3	5	US-08-171-382-4	1	
977	2.3	5	US-08-309-420-4	1	

978	5	2.3	454	1	US-08-309-419-4	Sequence 4,	Appli
979	5	2.3	454	2	US-09-014-969-9	Sequence 9,	Appli
980	5	2.3	454	4	PCT-US95-11856-4	Sequence 4,	Appli
981	5	2.3	454	4	PCT-US95-11878-4	Sequence 4,	Appli
982	5	2.3	455	2	US-08-870-827-3	Sequence 3,	Appli
983	5	2.3	457	2	US-09-057-969-5	Sequence 5,	Appli
984	5	2.3	460	1	US-08-403-866-3	Sequence 1,	Appli
985	5	2.3	462	2	US-08-865-597A-2	Sequence 2,	Appli
986	5	2.3	465	2	US-08-114-555A-2	Sequence 2,	Appli
987	5	2.3	465	2	US-08-833-678A-2	Sequence 2,	Appli
988	5	2.3	465	4	PCT-US93-08386-5	Sequence 5,	Appli
989	5	2.3	466	2	US-08-836-791-9	Sequence 9,	Appli
990	5	2.3	466	2	US-08-486-839-4	Sequence 4,	Appli
991	5	2.3	467	3	US-08-701-582D-2	Sequence 2,	Appli
992	5	2.3	467	3	US-08-701-582D-4	Sequence 4,	Appli
993	5	2.3	470	2	US-08-377-440A-1	Sequence 1,	Appli
994	5	2.3	471	2	US-08-466-583-2	Sequence 2,	Appli
995	5	2.3	471	4	PCT-US95-07820-2	Sequence 2,	Appli
996	5	2.3	473	2	PCT-US93-08386-9	Sequence 9,	Appli
997	5	2.3	475	1	US-07-686-591-4	Sequence 4,	Appli
998	5	2.3	475	1	US-07-970-715-4	Sequence 4,	Appli
999	5	2.3	477	1	US-07-791-936A-2	Sequence 2,	Appli
1000	5	2.3	477	1	US-08-402-217A-3	Sequence 3,	Appli

ALIGNMENTS

RESULT 1
US-08-225-488-7
; Sequence 7, Application US/08225488
; Patent No. 5846802
; GENERAL INFORMATION:
; APPLICANT: Buxton, Frank
; APPLICANT: Hinnen, Albert
; APPLICANT: Visser, Jacob
; TITLE OF INVENTION: No. 5846802el Fungal Protease
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,488
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/047,214
; FILING DATE: 13-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprull, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 4-19055/A/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-225-488-7

Query Match   3.2%;   Score 7;   DB 2;   Length 416;
Best Local Similarity 100.0%;   Pred. No. 34;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   139 LPAGIER 145
          |||||||
Db   75 LPAGIER 81

RESULT 2
US-08-769-309A-21
; Sequence 21, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 571890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-769-309A-21

Query Match   2.7%;   Score 6;   DB 1;   Length 22;
Best Local Similarity 100.0%;   Pred. No. 23;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   15 KTSFRK 20
          ||||||
Db   15 KTSFRK 20

RESULT 3
US-08-449-639-3
; Sequence 3, Application US/08449639
; Patent No. 5770191
; GENERAL INFORMATION:
; APPLICANT: Johnson, Howard M.
; APPLICANT: Szente, Brian E.
; TITLE OF INVENTION: C-Terminal Peptides of Interferon-Gamma
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1

```
            CITY: Gainesville
            STATE: FL
            COUNTRY: USA
            ZIP: 32606
        COMPUTER READABLE FORM:
            MEDIUM TYPE: Floppy disk
            COMPUTER: IBM PC compatible
            OPERATING SYSTEM: PC-DOS/MS-DOS
            SOFTWARE: PatentIn Release #1.0, Version #1.30
        CURRENT APPLICATION DATA:
            APPLICATION NUMBER: US/08/449,639
            FILING DATE:
            CLASSIFICATION: 530
        ATTORNEY/AGENT INFORMATION:
            NAME: Saliwanchik, David R.
            REGISTRATION NUMBER: 31,794
            REFERENCE/DOCKET NUMBER: UF150
        TELECOMMUNICATION INFORMATION:
            TELEPHONE: 904-375-8100
            TELEFAX: 904-375-8100
    INFORMATION FOR SEQ ID NO: 3:
        SEQUENCE CHARACTERISTICS:
            LENGTH: 39 amino acids
            TYPE: amino acid
            STRANDEDNESS: single
            TOPOLOGY: linear
        MOLECULE TYPE: peptide
US-08-449-639-3

Query Match      2.7%;    Score 6; DB 1;    Length 39;
Best Local Similarity    100.0%;    Pred. No. 38;
Matches   6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    20 KHRVER 25
         ||||||
Db    24 KHRVER 29

RESULT 4
US-08-902-623-43
; Sequence 43, Application US/08902623
; Patent No. 5922545
  GENERAL INFORMATION:
    APPLICANT: MATTHEAKIS, LARRY C.
    APPLICANT: DOWER, WILLIAM J.
    TITLE OF INVENTION: IN VITRO PEPTIDE AND ANTIBODY DISPLAY
    TITLE OF INVENTION: LIBRARIES
    NUMBER OF SEQUENCES: 75
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
        STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
        CITY: SAN FRANCISCO
        STATE: CALIFORNIA
        COUNTRY: U.S.A.
        ZIP: 94111-3834
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/902,623
        FILING DATE:
        CLASSIFICATION:
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/586,176
        FILING DATE: 17-JAN-1996
        APPLICATION NUMBER: US 08/300,262
        FILING DATE: 02-SEP-1994
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/144,775
        FILING DATE:. US 29-OCT-1993
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: PCT/US94/12206
        FILING DATE: US 25-OCT-1994
    ATTORNEY/AGENT INFORMATION:
        NAME: DUNN, TRACY J.
        REGISTRATION NUMBER: 34,587
        REFERENCE/DOCKET NUMBER: 16528X-003330US
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 326-2400
        TELEFAX: (415) 326-2422
  INFORMATION FOR SEQ ID NO: 43:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 93 amino acids
        TYPE: amino acid
        STRANDEDNESS: unknown
        TOPOLOGY: unknown
US-08-902-623-43

Query Match      2.7%;    Score 6; DB 2;    Length 93;
Best Local Similarity    100.0%;    Pred. No. 84;
Matches   6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    29 RQATRR 34
         ||||||
Db    83 RQATRR 88

RESULT 5
US-08-215-805A-82
; Sequence 82, Application US/08215805A
; Patent No. 5559008
  GENERAL INFORMATION:
    APPLICANT: Chang, Yung-Fu
    TITLE OF INVENTION: LEUKOTOXIN GENE FROM PASTEURELLA
    TITLE OF INVENTION: SUIS
    NUMBER OF SEQUENCES: 84
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Nixon, Hargrave, Devans & Doyle
        STREET: Clinton Square, P.O. Box 1051
        CITY: Rochester
        STATE: New York
        COUNTRY: USA
        ZIP: 14603
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/215,805A
        FILING DATE: 22-MAR-1994
        CLASSIFICATION: 435
    ATTORNEY/AGENT INFORMATION:
        NAME: Timian, Susan J
        REGISTRATION NUMBER: 34,103
        REFERENCE/DOCKET NUMBER: 19603/61 (D-1329A)
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (716) 263-1636
        TELEFAX: (716) 263-1600
  INFORMATION FOR SEQ ID NO: 82:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 165 amino acids
        TYPE: amino acid
        STRANDEDNESS:
        TOPOLOGY: linear
    MOLECULE TYPE: protein
    ORIGINAL SOURCE:
        ORGANISM: Pasteurella suis
        STRAIN: 5943
    IMMEDIATE SOURCE:
        LIBRARY: P. suis DNA in Bacteriophage lambda-dash
        CLONE: (Lambda)yfc33-37
```

US-08-215-805A-82

Query Match   2.7%; Score 6; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   105 KRTGKI 110
       ||||||
Db   124 KRTGKI 129

RESULT 6
US-08-933-750C-28
; Sequence 28, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
;   APPLICANT: Lal, Preeti
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Bandman, Olga
;   APPLICANT: Shah, Purvi
;   APPLICANT: Au-Young, Janice
;   APPLICANT: Yue, Henry
;   APPLICANT: Guegler, Karl J.
;   APPLICANT: Corley, Neil C.
;   TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
;   NUMBER OF SEQUENCES: 98
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/933,750C
;   FILING DATE: September 23, 1997
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0356 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
;   TELEX:
;   INFORMATION FOR SEQ ID NO: 28:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 179 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   LIBRARY: LNODNOT03
;   CLONE: 1577239
US-08-933-750C-28

Query Match   2.7%; Score 6; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   138 TLPAGI 143
       ||||||
Db   126 TLPAGI 131

RESULT 7
US-08-622-352A-3
; Sequence 3, Application US/08622352A
; Patent No. 5824546
; GENERAL INFORMATION:
;   APPLICANT: Bishai, William R.
;   APPLICANT: DeMaio, James
;   TITLE OF INVENTION: REGULATION OF A SIGMA FACTOR
;   TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
;   STREET: 1100 New York Avenue, N.W.
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20005-3918
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Word Perfect
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/622,352A
;   FILING DATE: 27-MAR-1996
;   CLASSIFICATION: 435
;   INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 242 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Mycobacterium tuberculosis
US-08-622-352A-3

Query Match   2.7%; Score 6; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   76 AAGRLD 81
      ||||||
Db   27 AAGRLD 32

RESULT 8
US-08-769-309A-3
; Sequence 3, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
;   APPLICANT: Scott, John D.,
;   APPLICANT: Nuert, Brian J.,
;   APPLICANT: Klauck, Theresa M.
;   TITLE OF INVENTION: Protein Binding Domains of Gravin
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower/233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:

```
;     APPLICATION NUMBER:  US/08/769,309A
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  No. 5741890and, Greta E.
;     REGISTRATION NUMBER:  35,302
;     REFERENCE/DOCKET NUMBER:  27866/33451
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  312-474-6300
;     TELEFAX:  312-474-0448
;     TELEX:  25-3856
; INFORMATION FOR SEQ ID NO:  3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  292 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
US-08-769-309A-3

Query Match            2.7%;  Score 6;  DB 1;  Length 292;
Best Local Similarity  100.0%;  Pred. No. 2.4e+02;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  15 KTSFRK 20
       ||||||
Db  45 KTSFRK 50

RESULT  9
US-08-375-962B-12
; Sequence 12, Application US/08375962B
; Patent No. 5731195
; GENERAL INFORMATION:
;   APPLICANT:  SIMON, ANDRAS; HELLMAN, ULF; WERNSTEDT,
;   APPLICANT:  CHRISTER; ERIKSSON, ULF.
;   TITLE OF INVENTION:  Isolated Nucleic Acid Molecule
;   TITLE OF INVENTION:  Which Codes for A 32 kDa Protein Having 11-CIS Retinol
;   TITLE OF INVENTION:  Dehydrogenase Activity, and Which Associates With P63,
;   TITLE OF INVENTION:  a Portion of a Retinol Binding Protein Receptor
;   NUMBER OF SEQUENCES:  14
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Felfe & Lynch
;     STREET:  805 Third Avenue
;     CITY:  New York City
;     STATE:  New York
;     ZIP:  10022
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette, 3.5 inch, 144 kb storage
;     COMPUTER:  IBM
;     OPERATING SYSTEM:  PC-DOS
;     SOFTWARE:  Wordperfect (ASCII standard)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/375,962B
;     FILING DATE:  20-January-1995
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/258,418
;     FILING DATE:  6-October-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Pasqualini, Patricia A.
;     REGISTRATION NUMBER:  34,894
;     REFERENCE/DOCKET NUMBER:  LUD 5372.1 CIP
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (212) 688-9200
;     TELEFAX:  (212) 838-3884
; INFORMATION FOR SEQ ID NO:  12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  327 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
```

```
;   FEATURE:
;     NAME/KEY:  Human estradiol 17-b dehydrogenase (hEDH)
US-08-375-962B-12

Query Match            2.7%;  Score 6;  DB 1;  Length 327;
Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   58 DEAGRS 63
        ||||||
Db  307 DEAGRS 312

RESULT  10
US-08-562-114B-12
; Sequence 12, Application US/08562114B
; Patent No. 5972646
; GENERAL INFORMATION:
;   APPLICANT:  ERIKSSON ET AL.
;   TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE WHICH CODES FOR A
;   TITLE OF INVENTION:  32 KDA PROTEIN HAVING 11-CIS RETINOL HYROGENASE
;   TITLE OF INVENTION:  ACTIVITY, AND WHICH ASSOCIATES WITH P63, A
;   TITLE OF INVENTION:  PORTION OF A RETINOL BINDING PROTEIN RECEPTOR
;   NUMBER OF SEQUENCES:  31
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Felfe & Lynch
;     STREET:  805 Third Avenue
;     CITY:  New York City
;     STATE:  New York
;     ZIP:  10022
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette, 3.5 inch, 144 kb storage
;     COMPUTER:  IBM
;     OPERATING SYSTEM:  PC-DOS
;     SOFTWARE:  Wordperfect 5.1 and ASCII
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/562,114B
;     FILING DATE:  22-No. 5972646ember-1995
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/375,962
;     FILING DATE:  20-January-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Kohli, Vineet
;     REGISTRATION NUMBER:  37,003
;     REFERENCE/DOCKET NUMBER:  LUD 5372.2 CIP
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (212) 688-9200
;     TELEFAX:  (212) 838-3884
; INFORMATION FOR SEQ ID NO:  12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  327 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;   FEATURE:
;     NAME/KEY:  Human estradiol 17-b dehydrogenase (hEDH)
US-08-562-114B-12

Query Match            2.7%;  Score 6;  DB 2;  Length 327;
Best Local Similarity  100.0%;  Pred. No. 2.6e+02;
Matches  6;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   58 DEAGRS 63
        ||||||
Db  307 DEAGRS 312

RESULT  11
US-08-855-714-3
; Sequence 3, Application US/08855714
```

Patent No. 5939075
GENERAL INFORMATION:
; APPLICANT: Houng, Huo-Shu H.
; APPLICANT: Warren, Richard L.
; TITLE OF INVENTION: MUTANTS OF BRUCELLA MELITENSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John Moran, Esq.
; STREET: HQ USAMRDC, Dept. of Army, Fort Detrick
; CITY: Frederick
; STATE: MD
; COUNTRY: US
; ZIP: 21702-5012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,714
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/334,129
; FILING DATE: 04-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: 08/143,692
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 619-2065
; TELEFAX: (301) 619-7714
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 339 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: both
;   TOPOLOGY: unknown
; MOLECULE TYPE:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
;   ORGANISM: BRUCELLA MELITENSIS
US-08-855-714-3

Query Match   2.7%;   Score 6;   DB 2;   Length 339;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   166 PTSWLK 171
          ||||||
Db   64 PTSWLK 69

RESULT 12
US-08-467-559B-9
; Sequence 9, Application US/08467559B
; Patent No. 5928890
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; TITLE OF INVENTION: HUMAN AMINE RECEPTOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
;   STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
;   CITY: WASHINGTON
;   STATE: DC
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 20005-3934
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/467,559B
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: STEFFE, ERIC K
;   REGISTRATION NUMBER: 36,688
;   REFERENCE/DOCKET NUMBER: 1488.0840000
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 365 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-467-559B-9

Query Match   2.7%;   Score 6;   DB 2;   Length 365;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   207 DSLANG 212
          ||||||
Db   259 DSLANG 264

RESULT 13
US-08-846-762-89
; Sequence 89, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 89
;   LENGTH: 403
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-08-846-762-89

Query Match   2.7%;   Score 6;   DB 2;   Length 403;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   49 PYDVLP 54
          ||||||
Db   98 PYDVLP 103

RESULT 14
US-09-031-485-15
; Sequence 15, Application US/09031485
; Patent No. 5824306
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND

; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,485
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,429
; FILING DATE: 24-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-031-485-15

Query Match
Best Local Similarity   2.7%;   Score 6; DB 2; Length 409;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   38 PQPTRV 43
           ||||||
Db   210 PQPTRV 215

RESULT 15
US-08-847-429A-15
; Sequence 15, Application US/08847429A
; Patent No. 5827692
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Blehm, E. Scot
; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,429A

; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-847-429A-15

Query Match
Best Local Similarity   2.7%;   Score 6; DB 2; Length 409;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   38 PQPTRV 43
           ||||||
Db   210 PQPTRV 215

Search completed: June 10, 2000, 11:31:29
Job time: 343 sec

us-09-252-691-7056.oligo.rai

Copyright   GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2000, 10:45:07 ; Search time 14.58 seconds
                                          (without alignments)
                                          892.716 Million cell updates/sec

Title: US-09-252-691-7056
Perfect score: 222
Sequence: 1 AIMRQLITPENTMTKTSFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table: OLIGO
                   Gapop 60.0 , Gapext 60.0

Searched: 168808 seqs, 58629743 residues

Word size :   0

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database :   PIR_63:*
         1: pir1:*
         2: pir2:*
         3: pir3:*
         4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	52	23.4	207	2	D64858	probable pseudouri
2	15	6.8	240	2	I64156	probable pseudouri
3	12	5.4	185	2	S74428	probable pseudouri
4	10	4.5	229	2	S45555	conserved hypothet
5	10	4.5	257	2	D75462	ribosomal large su
6	8	3.6	341	2	T19185	hypothetical prote
7	8	3.6	448	2	A71347	hypothetical prote
8	8	3.6	555	2	B41492	58K antigen - Rick
9	8	3.6	577	2	B75621	sensor histidine k
10	7	3.2	241	2	JE0169	dnaJ heat shock pr
11	7	3.2	250	2	A43662	replicating protei
12	7	3.2	288	2	S41382	pobR protein - Pse
13	7	3.2	312	2	S74669	hypothetical prote
14	7	3.2	406	2	A71965	hypothetical prote
15	7	3.2	444	2	T35794	probable 4-aminobu
16	7	3.2	514	2	B64634	hypothetical prote
17	7	3.2	514	2	E71880	hypothetical prote
18	7	3.2	515	2	S57942	protein disulfide-
19	7	3.2	532	1	QFPGM	neurofilament trip
20	7	3.2	534	2	S56785	TCP1-related prote
21	7	3.2	546	2	S48086	t-complex-type mol
22	7	3.2	604	2	S58643	probable transcrip
23	7	3.2	707	1	A53796	gelatinase B (EC 3
24	7	3.2	764	2	S49849	aconitate hydratas
25	7	3.2	845	2	A45669	neurofilament trip
26	7	3.2	849	2	S00030	neurofilament trip
27	7	3.2	898	2	T10101	aconitate hydratas
28	7	3.2	916	2	A27864	neurofilament trip
29	7	3.2	990	1	TQEC26	transposase - Esch
30	7	3.2	1112	3	T30202	probable chitin sy
31	7	3.2	1148	1	G64855	transcription/repa
32	7	3.2	1886	2	S04921	nuclear pore prote
33	6	2.7	43	2	A55024	cholic acid transp
34	6	2.7	68	2	B75336	hypothetical prote
35	6	2.7	69	2	E69385	hypothetical prote
36	6	2.7	71	2	T03198	hypothetical prote
37	6	2.7	89	1	WMVZRF	10K cell fusion pr
38	6	2.7	90	2	C56272	prgs protein - Ent
39	6	2.7	90	2	E41662	probable pheromone
40	6	2.7	93	2	T35008	probable ROK famil
41	6	2.7	94	2	JC4063	integration host f
42	6	2.7	97	2	S07266	hypothetical prote
43	6	2.7	103	2	IKECV	colicin v - Escher
44	6	2.7	106	2	S34814	hypothetical prote
45	6	2.7	111	2	B75351	conserved hypothet
46	6	2.7	117	2	T02603	hypothetical prote
47	6	2.7	119	2	T28757	hypothetical prote
48	6	2.7	133	2	S28187	interleukin-4 - pi
49	6	2.7	135	2	S34952	histone H1 - Euplo
50	6	2.7	138	2	T16185	hypothetical prote
51	6	2.7	141	2	S67149	hypothetical prote
52	6	2.7	142	2	A47169	membrane-bound mul
53	6	2.7	145	2	T36091	probable maturase-
54	6	2.7	146	2	I39503	gene aac(6')-Ih pr
55	6	2.7	148	2	S64796	hypothetical prote
56	6	2.7	150	2	C37152	hypothetical prote
57	6	2.7	154	2	S40522	Huntington's disea
58	6	2.7	158	2	D36138	ureE protein - Kle
59	6	2.7	171	2	A45562	merozoite surface
60	6	2.7	173	2	T12927	endodeoxyribonucle
61	6	2.7	173	2	S77292	hypothetical proce
62	6	2.7	173	2	A49624	polyprotein - chic
63	6	2.7	174	2	I77413	gamma-D-crystallin
64	6	2.7	174	2	S41713	plu protein - frui
65	6	2.7	180	2	T41322	probable succinate
66	6	2.7	190	2	T29068	hypothetical prote
67	6	2.7	191	2	B26602	apolipoprotein 2 p
68	6	2.7	192	2	A28571	beta-glucosidase (
69	6	2.7	193	2	F64158	hypothetical prote
70	6	2.7	194	2	T31949	hypothetical prote
71	6	2.7	205	1	CYBOB	beta-crystallin B2
72	6	2.7	209	2	I64172	hypothetical prote
73	6	2.7	209	2	T34812	extragenic suppres
74	6	2.7	216	2	S61701	hypothetical prote
75	6	2.7	217	2	WJHU2C	homeotic protein H
76	6	2.7	217	1	WJMSX2	homeotic protein H
77	6	2.7	223	2	T30440	hypothetical prote
78	6	2.7	225	2	S36621	probable drought-1
79	6	2.7	228	2	H70592	probable mtrA prot
80	6	2.7	228	2	T36032	hypothetical prote
81	6	2.7	231	2	F64987	hypothetical 25.9
82	6	2.7	233	2	E72301	hypothetical prote
83	6	2.7	235	1	B41326	nitrile hydratase
84	6	2.7	235	2	F70127	probable glucose-6
85	6	2.7	238	2	S75830	hypothetical prote
86	6	2.7	239	2	S59050	tryptophan synthas
87	6	2.7	240	2	D72415	zinc ABC transport
88	6	2.7	242	2	H70980	probable rsbW prot
89	6	2.7	242	2	T37107	hypothetical prote
90	6	2.7	247	2	F69726	pseudouridylate sy
91	6	2.7	249	2	S76104	hypothetical prote
92	6	2.7	249	2	G75583	hypothetical prote
93	6	2.7	250	2	A69951	N-acetylmuramoyl-L
94	6	2.7	250	2	S44957	lmbJ protein - Str
95	6	2.7	253	2	D72569	hypothetical prote
96	6	2.7	254	1	BVBY53	phosphomannomutase
97	6	2.7	254	2	B70235	hypothetical prote
98	6	2.7	258	2	C70968	probable ribD - My
99	6	2.7	260	1	JKECT	lysine-arginine-or
100	6	2.7	260	1	B64439	hypothetical prote
101	6	2.7	261	2	T35708	hydrolase - Strept
102	6	2.7	262	2	S63648	H+-transporting AT

Left column group

No.			ID	n	Accession	Description
103	2.7	6	266	2	S06462	beta-lactamase (EC
104	2.7	6	266	2	S33163	beta-lactamase (EC
105	2.7	6	269	2	T36910	hypothetical prote
106	2.7	6	270	2	T22151	hypothetical prote
107	2.7	6	271	2	A72200	septum site-determ
108	2.7	6	272	2	A26936	NAD+ synthase (EC
109	2.7	6	274	2	H70900	hypothetical prote
110	2.7	6	276	2	T05778	hypothetical prote
111	2.7	6	277	2	S38503	beta-lactamase (EC
112	2.7	6	277	2	S38671	beta-lactamase (EC
113	2.7	6	281	2	E72395	sugar ABC transpor
114	2.7	6	281	2	G71204	hypothetical prote
115	2.7	6	284	2	E75018	endonuclease iv re
116	2.7	6	284	2	S62931	probable membrane
117	2.7	6	285	2	S62955	probable membrane
118	2.7	6	288	2	T14717	yhgA protein homol
119	2.7	6	288	2	S44256	fructokinase (EC 2
120	2.7	6	290	2	A71148	hypothetical prote
121	2.7	6	293	1	H70530	probable extrageni
122	2.7	6	294	2	S72970	tyrosine--tRNA lig
123	2.7	6	298	2	T33338	probable transcrip
124	2.7	6	298	2	T18046	probable DNA ligas
125	2.7	6	300	2	B75612	phosphate ABC tran
126	2.7	6	300	2	H70344	6-phosphogluconate
127	2.7	6	301	2	T13535	N-acetylmuramoyl-L
128	2.7	6	303	2	D44355	CD44 glycoprotein
129	2.7	6	306	2	A75301	SudD-related prote
130	2.7	6	307	2	A48118	major epidermal ca
131	2.7	6	312	2	T34651	probable transmemb
132	2.7	6	313	2	E69045	8-oxoguanine DNA g
133	2.7	6	313	2	XWNC	ADP,ATP carrier pr
134	2.7	6	313	2	S41385	hypothetical yadD
135	2.7	6	318	2	T35177	iron-sulfur oxidor
136	2.7	6	319	1	D64417	transformation sen
137	2.7	6	319	2	S61293	transcription init
138	2.7	6	321	2	A69756	adhesion protein h
139	2.7	6	323	2	H65173	hypothetical trans
140	2.7	6	323	2	C70653	probable prephenat
141	2.7	6	324	2	A36444	rho-crystallin - b
142	2.7	6	324	2	F75532	MutT/nudix family
143	2.7	6	328	2	JC4280	rho-crystallin - b
144	2.7	6	328	2	T27302	hypothetical prote
145	2.7	6	328	1	DEHUE7	estradiol 17beta-d
146	2.7	6	338	2	T14989	probable transposa
147	2.7	6	338	2	T19815	hypothetical prote
148	2.7	6	338	2	S58158	hypothetical prote
149	2.7	6	340	2	S38562	hypothetical prote
150	2.7	6	342	2	S46819	hypothetical prote
151	2.7	6	344	2	T15458	hypothetical prote
152	2.7	6	346	2	T40774	probable histone p
153	2.7	6	351	3	C71603	pseudouridine synt
154	2.7	6	354	2	T40327	probable pseudouri
155	2.7	6	355	2	G70741	hypothetical prote
156	2.7	6	356	2	C71372	conserved hypothet
157	2.7	6	357	2	T21604	hypothetical prote
158	2.7	6	359	2	A34365	holocytochrome-c s
159	2.7	6	360	2	T09163	probable peroxidas
160	2.7	6	360	1	P1WLR1	L1 protein - rhesu
161	2.7	6	362	2	S21057	recF protein - Sal
162	2.7	6	363	2	S39889	SMR1 protein - Pod
163	2.7	6	365	2	T19312	hypothetical prote
164	2.7	6	368	2	RQECF	recF protein - Esc
165	2.7	6	371	1	S27788	neutrophil oxidase
166	2.7	6	371	1	A25732	inhibin alpha chai
167	2.7	6	375	2	E70892	hypothetical prote
168	2.7	6	375	1	JQ0735	recF protein - Pro
169	2.7	6		2	I39726	mannopine biosynth
170	2.7	6		1	JQ0880	NS5 protein - hepa
171	2.7	6		2	S52202	X-Pro dipeptidase
172	2.7	6		2	S60121	mitogen-activated
173	2.7	6		2	H72045	flagellar motor sw
174	2.7	6		2	I64106	recF protein - Hae
175	2.7	6		2	A75264	hypothetical prote

Right column group

Description	Accession	n	No.		
hypothetical prote	D70957	2	377	2.7	6
zinc metalloprotei	T00643	2	377	2.7	6
probable sarcosine	B71184	2	377	2.7	6
sarcosine oxidase,	F75144	2	379	2.7	6
hypothetical prote	F72352	2	383	2.7	6
probable S2P metal	G77777	2	383	2.7	6
thyroid hormone re	S41497	2	386	2.7	6
probable ABC trans	T35854	2	386	2.7	6
hypothetical prote,	H64427	2	388	2.7	6
probable regulator	C70754	2	388	2.7	6
hypothetical prote	G72491	2	390	2.7	6
hypothetical prote	S74688	2	391	2.7	6
hypothetical prote	T32429	2	391	2.7	3
cytochrome P450 hy	T17486	2	391	2.7	6
probable acyl-CoA	E72549	2	393	2.7	6
hypothetical prote	G75587	2	395	2.7	6
isp7 protein - fis	S45496	2	397	2.7	6
hypothetical prote	B70752	2	398	2.7	6
polymerase-associa	A43685	1	399	2.7	6
actin-binding prot	S32565	2	402	2.7	6
hypothetical prote	H64142	1	403	2.7	6
glycosyltransferas	S70676	2	403	2.7	6
hypothetical ROK f	T34923	2	403	2.7	6
procyclin-associat	A48447	2	405	2.7	6
cytochrome P450 hy	T17487	3	406	2.7	6
hypothetical prote	F69805	2	407	2.7	6
hypothetical prote	F70696	2	407	2.7	6
pseudouridine synt	T39326	2	407	2.7	6
hypothetical prote	G69874	2	408	2.7	6
methylaspartate am	B44285	2	413	2.7	6
hypothetical prote	T26572	2	413	2.7	6
glycine hydroxymet	JQ1016	1	414	2.7	6
CD44 protein - mou	S30397	2	414	2.7	6
phosphoglycerate k	KIFFPG	1	415	2.7	6
hypothetical prote	G64647	2	416	2.7	6
hypothetical prote	D71936	2	416	2.7	6
hypothetical prote	D72382	2	425	2.7	6
cyclin cycl - Arab	T04743	2	428	2.7	6
fibrinogen-like pr	I56934	2	432	2.7	6
cytotoxic T-lympho	A27447	2	432	2.7	6
hypothetical prote	T21989	2	433	2.7	6
probable electron	D72472	2	436	2.7	6
protein kinase hom	T04655	2	438	2.7	6
fibrinogen-like pr	I37391	2	439	2.7	6
sensory transducti	E69102	2	444	2.7	6
nickel resistance	S77010	1	445	2.7	6
hypothetical prote	B64894	2	447	2.7	6
centrosomin B - mo	S52391	2	447	2.7	6
NADH oxidase - Del	B75451	2	449	2.7	6
yaeL protein - Esc	H64741	2	450	2.7	6
conserved hypothet	C70210	2	452	2.7	6
hypothetical prote	C70728	2	452	2.7	6
hypothetical prote	T32300	2	456	2.7	6
dihydrolipoamide d	D69962	2	457	2.7	6
probable mercuric	B72625	2	461	2.7	6
beta-1-adrenergic	S12591	2	464	2.7	6
dopamine D1C recep	S15661	2	465	2.7	6
beta-1-adrenergic	S36794	2	466	2.7	6
probable fumarate	H70896	2	474	2.7	6
hypothetical prote	T27147	2	485	2.7	6
aldehyde dehydroge	I39769	1	488	2.7	6
probable penicilli	T36716	2	490	2.7	6
probable serine-ty	B71279	2	490	2.7	6
DEAD-box RNA helic	T34592	2	498	2.7	6
L1 protein - human	P1WLL3	1	499	2.7	6
L1 protein - human	P1WL33	2	499	2.7	6
L1 protein - human	S36531	1	499	2.7	6
L1 protein - human	P1WL6	1	500	2.7	6
cytochrome P450 ib	A40938	2	501	2.7	6
L1 protein - human	P1WL11	1	501	2.7	6
cytochrome P450 ho	T04735	2	501	2.7	6
membrane-bound pro	JC5739	2	501	2.7	6
L1 protein - pygmy	P1WLC1	1	502	2.7	6

#			code	#		description		#	code	#			description
249	2.7	6	P1WL42	502	1	L1 protein - human		322	T18063	649	2	2.7 6	DNA binding protei
250	2.7	6	S36549	503	2	L1 protein - human		323	JC4610	651	2	2.7 6	dnaK-type molecula
251	2.7	6	P1WLPR	505	1	L1 protein - human		324	S64073	657	2	2.7 6	hypothetical prote
252	2.7	6	P1WL39	505	1	L1 protein - human		325	D71351	657	2	2.7 6	probable primosoma
253	2.7	6	JC5054	507	2	high affinity nitr		326	A53668	661	2	2.7 6	glycoprotein gp100
254	2.7	6	S36508	508	2	L1 protein - human		327	I38400	662	1	2.7 6	melanoma-associate
255	2.7	6	T31630	508	1	hypothetical prote		328	E69475	663	2	2.7 6	DNA topoisomerase
256	2.7	6	JC5055	509	2	high affinity nitr		329	A30185	665	2	2.7 6	nuclear protein nu
257	2.7	6	S15620	510	2	L1 protein - human		330	A41234	668	2	2.7 6	melanocyte-specif1
258	2.7	6	S15627	510	2	L1 protein - human		331	T38181	669	2	2.7 6	glutamate--cystein
259	2.7	6	T36306	511	2	probable anthranil		332	S61462	672	2	2.7 6	p83/100 protein -
260	2.7	6	T30924	512	2	hypothetical prote		333	S61463	672	2	2.7 6	p83/100 protein -
261	2.7	6	B72752	514	2	hypothetical prote		334	T19495	674	2	2.7 6	hypothetical prote
262	2.7	6	T35371	514	2	probable Na+/H+ an		335	S09078	677	1	2.7 6	chromogranin B pre
263	2.7	6	T25534	515	2	hypothetical prote		336	JS0178	677	2	2.7 6	protein kinase YKR
264	2.7	6	CSBYP	515	2	catalase (EC 1.11.		337	E71282	679	2	2.7 6	probable NH(3)-dep
265	2.7	6	I49154	517	2	calcitonin recepto		338	S76658	684	2	2.7 6	hypothetical prote
266	2.7	6	T25615	520	2	hypothetical prote		339	JC6331	685	2	2.7 6	rho-type guanine e
267	2.7	6	T06625	522	1	hypothetical prote		340	T32980	692	2	2.7 6	hypothetical prote
268	2.7	6	S34190	524	2	sulfite reductase		341	S61464	693	2	2.7 6	p83/100 protein -
269	2.7	6	P1WL58	526	2	L1 protein - human		342	I40090	693	2	2.7 6	p93 protein - Lyme
270	2.7	6	D72107	529	2	oligopeptide bindi		343	I40141	693	2	2.7 6	p93 protein - Lyme
271	2.7	6	S36578	531	2	L1 protein - human		344	A69222	698	2	2.7 6	hypothetical prote
272	2.7	6	S36537	532	2	L1 protein - human		345	T32840	698	2	2.7 6	hypothetical prote
273	2.7	6	S36554	533	2	L1 protein (altern		346	T16176	709	2	2.7 6	hypothetical prote
274	2.7	6	S18539	534	2	actVA-1 protein -		347	F75440	711	2	2.7 6	methylmalonyl-CoA
275	2.7	6	S36583	537	2	L1 protein - human		348	S52205	722	2	2.7 6	mesI protein - Leu
276	2.7	6	S54770	539	2	secreted acid phos		349	T38220	728	2	2.7 6	hypothetical prote
277	2.7	6	S36566	544	2	L1 protein - human		350	T00343	738	2	2.7 6	hypothetical prote
278	2.7	6	S15695	545	2	4-coumarate--CoA l		351	T06017	746	2	2.7 6	subtilisin-like pr
279	2.7	6	S24389	547	2	hybrid cluster [4F		352	S27985	746	2	2.7 6	cellulase (EC 3.2.
280	2.7	6	F69376	548	2	dihydroxy-acid deh		353	T01441	760	2	2.7 6	hypothetical prote
281	2.7	6	E64395	548	2	hybrid cluster [4F		354	JQ1638	775	2	2.7 6	outer layer protei
282	2.7	6	S56152	550	2	H+-transporting AT		355	VPXRWF	775	1	2.7 6	outer layer protei
283	2.7	6	C75557	552	2	hypothetical prote		356	S50054	779	2	2.7 6	chloride channel p
284	2.7	6	T25195	553	1	hybrid cluster [4F		357	H69980	786	2	2.7 6	single-strand DNA-
285	2.7	6	S29861	554	2	glutamine--tRNA li		358	T25095	790	2	2.7 6	hypothetical prote
286	2.7	6	SYECQT	554	2	probable small nuc		359	T05247	797	2	2.7 6	methionine--tRNA l
287	2.7	6	T02445	554	2	hypothetical prote		360	T21315	802	2	2.7 6	hypothetical prote
288	2.7	6	S67452	557	2	hypothetical prote		361	I40508	810	1	2.7 6	endopeptidase Clp
289	2.7	6	T38067	566	2	viron protein - hu		362	S51451	814	2	2.7 6	probable membrane
290	2.7	6	T41952	568	2	nuclear scaffold-1		363	S48911	823	2	2.7 6	hypothetical prote
291	2.7	6	T11581	569	2	heat shock protein		364	E72305	835	2	2.7 6	hypothetical prote
292	2.7	6	T38942	573	2	L1 protein - human		365	T01357	840	2	2.7 6	hypothetical prote
293	2.7	6	P1WL18	573	1	C conserved hypoth		366	E71808	841	2	2.7 6	probable iron (III
294	2.7	6	A69511	575	2	catalase (EC 1.11.		367	H64694	842	2	2.7 6	iron(III) dicitrat
295	2.7	6	CSBYT	577	2	hypothetical prote		368	A48228	847	2	2.7 6	beta-N-acetylhexos
296	2.7	6	T19880	582	2	single-stranded-DN		369	G70483	856	2	2.7 6	pyruvate,water dik
297	2.7	6	F64110	584	2	single-stranded DN		370	S49046	859	2	2.7 6	luxQ protein - Vib
298	2.7	6	D65073	587	2	hypothetical prote		371	T05941	862	2	2.7 6	lipoxygenase (EC 1
299	2.7	6	C36329	587	2	s-glicerin precurso		372	S37088	863	1	2.7 6	phycobilisome anch
300	2.7	6	A53583	590	2	hypothetical prote		373	T05945	864	2	2.7 6	lipoxygenase (EC 1
301	2.7	6	I50419	591	2	fiber protein - mo		374	S23454	865	1	2.7 6	lipoxygenase (EC 1
302	2.7	6	C70893	594	2	hypothetical prote		375	A28443	869	2	2.7 6	phosphatidylethano
303	2.7	6	F71411	598	2	glutamic acid-rich		376	S72541	877	1	2.7 6	nitrate reductase
304	2.7	6	A40437	599	2	profilaggrin - hum		377	A40091	878	2	2.7 6	interleukin-3 rece
305	2.7	6	A45135	600	2	L1 protein (altern		378	S64755	879	2	2.7 6	hypothetical prote
306	2.7	6	S36502	602	2	probable U3 small		379	A35782	896	1	2.7 6	cytokine receptor
307	2.7	6	T38241	606	2	hypothetical prote		380	I56563	896	1	2.7 6	interleukin-3 rece
308	2.7	6	G71682	611	2	GTP-binding protei		381	A39255	897	2	2.7 6	cytokine receptor
309	2.7	6	A53583	606	2	geranylgeranyl-dip		382	S62517	906	2	2.7 6	hypothetical prote
310	2.7	6	T08180	606	2	PF20 protein, micr		383	T39123	906	2	2.7 6	hypothetical prote
311	2.7	6	T21747	611	2	hypothetical prote		384	S64155	910	2	2.7 6	hypothetical prote
312	2.7	6	ERADFM	613	1	fiber protein - mo		385	S38167	910	2	2.7 6	hypothetical prote
313	2.7	6	S38802	614	2	hypothetical prote		386	S39295	921	2	2.7 6	beta-adaptin 1 - f
314	2.7	6	F37257	622	2	hypothetical prote		387	I38100	946	2	2.7 6	p115 - human
315	2.7	6	S59234	623	2	glutamate--cystein		388	KIZMPO	947	1	2.7 6	pyruvate,orthophos
316	2.7	6	S53871	626	2	Pmel 17 protein		389	T02077	947	2	2.7 6	pyruvate,orthophos
317	2.7	6	T14776	629	2	hypothetical prote		390	S55478	949	1	2.7 6	pyruvate,orthophos
318	2.7	6	A45949	636	2	merozoite surface		391	D75377	951	2	2.7 6	probable proteinas
319	2.7	6	F70444	638	2	threonine--tRNA li		392	S12894	953	2	2.7 6	pyruvate,orthophos
320	2.7	6	S67656	640	2	hypothetical prote		393	S56650	953	2	2.7 6	pyruvate,orthophos
321	2.7	6	A55093	646	1	fatty acid transpo		394	S56649	955	2	2.7 6	pyruvate,orthophos

Left column

Index			Accession			Description
395	2.7	6	S53297	1	956	pyruvate,orthophos
396	2.7	6	A71420	2	960	pyruvate,orthophos
397	2.7	6	JQ1634	1	971	outer capsid prote
398	2.7	6	JH0109	2	979	glycoprotein 14 pr
399	2.7	6	VGBE2H	1	980	glycoprotein B pre
400	2.7	6	VGBEC6	2	980	glycoprotein B pre
401	2.7	6	T03784	1	990	probable receptor
402	2.7	6	T26746	2	1003	hypothetical prote
403	2.7	6	C71139	1	1003	hypothetical prote
404	2.7	6	S38003	2	1008	translation elonga
405	2.7	6	T31109	2	1014	myosin III - Atlan
406	2.7	6	H75423	1	1021	hypothetical prote
407	2.7	6	T38131	1	1033	hypothetical prote
408	2.7	6	C64418	2	1039	isoleucine--tRNA l
409	2.7	6	T22762	2	1049	hypothetical prote
410	2.7	6	T14189	1	1054	hypothetical prote
411	2.7	6	S18211	2	1063	hypothetical prote
412	2.7	6	T06312	1	1067	hypothetical prote
413	2.7	6	S51899	2	1081	probable protein k
414	2.7	6	T41188	2	1108	probable ubiquitin
415	2.7	6	S20497	2	1123	phytochrome A - po
416	2.7	6	A43736	2	1174	creatine kinase (E
417	2.7	6	S31145	2	1179	DNA-directed RNA p
418	2.7	6	T30160	1	1234	hypothetical prote
419	2.7	6	T42020	2	1239	class IV chitin sy
420	2.7	6	JC5209	2	1240	insulin receptor s
421	2.7	6	JS0670	1	1242	insulin receptor s
422	2.7	6	JC5615	2	1243	membrane-associate
423	2.7	6	T21269	2	1270	hypothetical prote
424	2.7	6	T37528	2	1288	probable snf2 fami
425	2.7	6	A31317	2	1290	1-phosphatidylinos
426	2.7	6	A36466	2	1290	1-phosphatidylinos
427	2.7	6	S00666	1	1291	1-phosphatidylinos
428	2.7	6	VCBES5	2	1330	major capsid prote
429	2.7	6	T13254	1	1350	nitric-oxide synth
430	2.7	6	T13887	2	1385	tlr protein - frui
431	2.7	6	T14158	1	1385	neurexin IV - mous
432	2.7	6	T13852	2	1389	gene wheeler prote
433	2.7	6	I50090	1	1395	carboxypeptidase g
434	2.7	6	S25997	2	1395	gene atpA intron 1
435	2.7	6	A60026	2	1475	cell communication
436	2.7	6	S28669	2	1481	amylopullulanase p
437	2.7	6	A47341	2	1481	amylopullulanase p
438	2.7	6	T13798	1	1494	hypothetical prote
439	2.7	6	S72250	2	1497	sex-determining tr
440	2.7	6	A40203	2	1515	4-alpha-glucanotra
441	2.7	6	T30518	2	1524	carbamoyl-phosphat
442	2.7	6	S38758	2	1555	amylo-1,6-glucosid
443	2.7	6	S53069	3	1562	probable membrane
444	2.7	6	T17411	2	1567	polyketide synthas
445	2.7	6	S11672	1	1571	ice nucleation pro
446	2.7	6	T00062	1	1592	hypothetical prote
447	2.7	6	T16055	2	1615	hypothetical prote
448	2.7	6	B49502	2	1684	protein-tyrosine-p
449	2.7	6	JW0057	1	1684	gravin - human
450	2.7	6	T14005	2	1710	phospholipase D (E
451	2.7	6	A39401	2	1726	merozoite surface
452	2.7	6	A45604	2	1751	major blood-stage
453	2.7	6	A49502	2	1767	protein-tyrosine-p
454	2.7	6	T19430	1	1820	hypothetical prote
455	2.7	6	T38781	2	1842	fatty acid synthas
456	2.7	6	A34092	1	1859	DNA-directed RNA p
457	2.7	6	T29959	2	1862	DNA-directed RNA p
458	2.7	6	T30938	2	2051	receptor tyrosine
459	2.7	6	BWASBE	2	2073	bimE protein - Eme
460	2.7	6	T13612	2	2118	hypothetical prote
461	2.7	6	GNNY2W	2	2205	genome polyprotein
462	2.7	6	GNNY4P	2	2206	genome polyprotein
463	2.7	6	GNNY27	1	2206	genome polyprotein
464	2.7	6	GNNY21	1	2206	genome polyprotein
465	2.7	6	S03822	1	2206	genome polyprotein
466	2.7	6	GNNY1P	1	2207	genome polyprotein
467	2.7	6	GNNY5P	1	2207	genome polyprotein

Right column

Index			Accession			Description
468	2.7	6	S09553	2	2207	genome polyprotein
469	2.7	6	GNNY2P	1	2209	genome polyprotein
470	2.7	6	GNNY3P	1	2209	genome polyprotein
471	2.7	6	A48548	1	2214	genome polyprotein
472	2.7	6	A35938	1	2248	profilaggrin - hum
473	2.7	6	A42548	1	2261	genome polyprotein
474	2.7	6	T37320	2	2514	hypothetical prote
475	2.7	6	T28675	2	2533	hypothetical prote
476	2.7	6	T28674	2	2533	ataxia telangiecta
477	2.7	6	T24588	2	2619	alpha-5ID-immobili
478	2.7	6	S59413	2	2628	alpha-5ID-immobili
479	2.7	6	T37919	2	2670	hypothetical prote
480	2.7	6	T30213	2	2809	probable membrane
481	2.7	6	C64474	2	2894	GCN1 homolog - fis
482	2.7	6	GNWVCH	1	3011	G-cadherin - sea u
483	2.7	6	S40770	2	3011	genome polyprotein
484	2.7	6	S46009	2	3092	polyprotein precur
485	2.7	6	A46068	2	3144	GTPase-activating
486	2.7	6	T30342	2	3161	Huntington disease
487	2.7	6	T17440	2	3163	protein HMWP1 - Ye
488	2.7	6	T17467	3	3413	probable polyketid
489	2.7	6	S23070	2	3573	rifamycin polyketi
490	2.7	6	T36410	2	3643	erythronolide synt
491	2.7	6	A53102	1	4543	probable peptide s
492	2.7	6	T31105	2	4919	alpha-2-macroglobu
493	2.7	6	T13931	2	6658	hypothetical prote
494	2.3	5	S36668	2	13	projectin - fruit
495	2.3	5	B60894	2	19	gamma crystallin I
496	2.3	5	JP0063	2	19	ribosomal protein
497	2.3	5	JP0064	2	21	methanol:N,N'-dime
498	2.3	5	S50118	2	25	ribosomal protein
499	2.3	5	PC2162	2	26	activating factor
500	2.3	5	D57001	2	28	angiotensin II rec
501	2.3	5	A43937	2	30	endo-1,4-beta-xyla
502	2.3	5	PL0189	2	30	endo-1,4-beta-xyla
503	2.3	5	T36103	2	31	Ig light chain - s
504	2.3	5	S21224	2	31	hypothetical prote
505	2.3	5	S27128	2	35	sperm protein EM5
506	2.3	5	PN0486	2	35	hypothetical prote
507	2.3	5	S62189	2	38	hypothetical prote
508	2.3	5	A60688	2	38	phytochrome - gard
509	2.3	5	T04074	2	40	transcription acti
510	2.3	5	S09338	2	40	ferredoxin [2Fe-2S
511	2.3	5	A58361	4	46	protein-tyrosine k
512	2.3	5	B47187	2	46	phosphoprotein pho
513	2.3	5	B26612	2	48	splenin - human
514	2.3	5	A26612	2	48	thymopoietin - hum
515	2.3	5	S27128	2	48	histone H1 - sea u
516	2.3	5	PN0486	2	48	cell division cycl
517	2.3	5	S35693	2	48	phosphatidylserine
518	2.3	5	TOBO1	1	49	cystic fibrosis tr
519	2.3	5	TOBO2	1	49	thymopoietin I - b
520	2.3	5	TOBO3	1	49	thymopoietin II -
521	2.3	5	Q9BPT3	1	49	thymopoietin III -
522	2.3	5	A54547	1	49	gene 19.5 protein
523	2.3	5	S72456	2	50	ribosomal protein
524	2.3	5	A27537	2	53	pachytene arrest p
525	2.3	5	T04382	2	53	rubredoxin - Chlor
526	2.3	5	T34769	2	54	ZEM5 protein - mai
527	2.3	5	S20835	2	55	nitrogen regulator
528	2.3	5	T03658	2	56	ribosomal protein
529	2.3	5	E33356	2	58	phosphoenolpyruvat
530	2.3	5	T06016	2	59	prothymosin alpha
531	2.3	5	JC1389	2	59	hypothetical prote
532	2.3	5	F75338	2	60	homeotic protein E
533	2.3	5	T04625	2	60	hypothetical prote
534	2.3	5	DNVPBF	2	61	hypothetical prote
535	2.3	5	B72657	2	61	DNA-binding protei
536	2.3	5	S25962	2	63	hypothetical prote
537	2.3	5	T12803	2	63	hypothetical prote
538	2.3	5	A48854	2	64	hypothetical prote
539	2.3	5	T07494	2	64	erythrocyte band 3
540	2.3	5	C36855	2	65	hypothetical prote / B3L protein - vari

Seq			N		Accession	Description
541	2.3	5	65	2	T28602	hypothetical prote
542	2.3	5	65	2	T15856	hypothetical Prote
543	2.3	5	65	2	A72172	H2L protein - vari
544	2.3	5	68	2	JW0050	G-protein gamma 7
545	2.3	5	68	2	A45128	GTP-binding regula
546	2.3	5	68	2	G72270	hypothetical prote
547	2.3	5	68	2	T29417	hypothetical prote
548	2.3	5	69	2	I56580	G-protein gamma su
549	2.3	5	69	2	A71330	hypothetical prote
550	2.3	5	70	2	B64091	ribosomal protein
551	2.3	5	70	2	S73211	ribosomal protein
552	2.3	5	70	2	T28640	Y4JM protein - Rhi
553	2.3	5	70	2	F69058	hypothetical prote
554	2.3	5	70	2	A05211	hypothetical prote
555	2.3	5	71	2	C75527	hypothetical prote
556	2.3	5	71	2	T15836	hypothetical prote
557	2.3	5	72	2	S78251	ribosomal protein
558	2.3	5	73	2	S77494	ribosomal protein
559	2.3	5	74	1	HHSY34	heat shock protein
560	2.3	5	74	2	F70526	hypothetical prote
561	2.3	5	75	2	T12080	low molecular welg
562	2.3	5	75	2	T01886	hypothetical prote
563	2.3	5	75	2	T26474	hypothetical prote
564	2.3	5	76	2	A46171	hypothetical prote
565	2.3	5	76	2	B65076	hypothetical prote
566	2.3	5	77	2	S30988	gene 43 protein -
567	2.3	5	77	2	T02773	y4dJ protein - Rhi
568	2.3	5	78	2	T30471	hypothetical prote
569	2.3	5	78	2	JQ1781	SalF6R protein - v
570	2.3	5	79	2	F75343	hypothetical prote
571	2.3	5	79	2	G69294	conserved hypothet
572	2.3	5	79	2	T03721	reverse transcript
573	2.3	5	80	1	OSHU7B	cytochrome-c oxida
574	2.3	5	80	2	A48284	intestinal trefoil
575	2.3	5	80	2	JQ1959	inorganic carbon t
576	2.3	5	81	1	WZBE49	gene 49 protein -
577	2.3	5	81	2	T36197	probable acyl carr
578	2.3	5	81	2	E72650	hypothetical prote
579	2.3	5	82	2	BVEC2E	cell growth regula
580	2.3	5	82	1	D60396	antigen 7H8/7 - Pl
581	2.3	5	82	2	E69453	hypothetical prote
582	2.3	5	83	2	T17809	hypothetical prote
583	2.3	5	83	2	T42102	hypothetical prote
584	2.3	5	83	2	G70765	ribosomal protein
585	2.3	5	84	2	S78266	probable ribosomal
586	2.3	5	84	2	T36353	late lactation pro
587	2.3	5	85	2	A28551	probable pyruvate,
588	2.3	5	87	2	C49917	retrovirus-related
589	2.3	5	87	2	B46200	pol polyprotein -
590	2.3	5	87	2	S32444	pol polyprotein -
591	2.3	5	87	2	S32445	pol polyprotein -
592	2.3	5	87	2	S32447	pol polyprotein -
593	2.3	5	87	2	S32449	M protein precurso
594	2.3	5	89	2	S60786	hypothetical prote
595	2.3	5	87	2	T22931	transcription acti
596	2.3	5	87	2	A35666	hypothetical prote
597	2.3	5	88	2	T00171	cytochrome-c oxida
598	2.3	5	88	1	OSBO7B	probable small sec
599	2.3	5	88	2	T36927	probable ribosomal
600	2.3	5	89	2	D71269	hypothetical prote
601	2.3	5	89	2	E75101	hypothetical prote
602	2.3	5	90	2	E70694	ribosomal protein
603	2.3	5	91	2	S72290	monocyte chemoattr
604	2.3	5	91	2	A46539	hypothetical prote
605	2.3	5	91	2	E70213	hypothetical prote
606	2.3	5	91	2	T03202	hypothetical prote
607	2.3	5	92	2	T22335	L-phenylalanine ox
608	2.3	5	92	2	S41662	hypothetical prote
609	2.3	5	93	2	S61391	nitrogen regulator
610	2.3	5	93	2	T35668	embryonic abundant
611	2.3	5	94	1	VUWTEM	integration host f
612	2.3	5	94	2	A64111	probable ferredoxi
613	2.3	5	94	2	I40210	probable membrane

Seq		Accession	N			Description
614	2	S44903	94	2.3	5	ZK652.3 protein -
615	3	T37725	94	2.3	5	hypothetical prote
616	1	E64043	95	2.3	5	citrate (pro-3S)-1
617	2	F64709	95	2.3	5	hypothetical prote
618	2	T26179	95	2.3	5	hypothetical prote
619	2	S65794	96	2.3	5	glutathione peroxi
620	2	A05308	96	2.3	5	tissue kallikrein
621	2	T07822	96	2.3	5	cystein proteinase
622	2	B46018	96	2.3	5	apolipoprotein C-I
623	2	JN0572	96	2.3	5	neutrophil chemo-a
624	2	S00119	96	2.3	5	hypothetical prote
625	2	C72309	96	2.3	5	hypothetical prote
626	1	FEBQ	97	2.3	5	ferredoxin [2Fe-2S
627	2	S69935	97	2.3	5	ferredoxin [2Fe-2S
628	2	B45271	97	2.3	5	MxiI - Shigella fl
629	2	T34765	97	2.3	5	small hydrophobic
630	2	S06808	98	2.3	5	keratin, feather (
631	2	S14883	98	2.3	5	hypothetical prote
632	2	F64317	99	2.3	5	hypothetical prote
633	2	C71198	99	2.3	5	hypothetical prote
634	2	T34614	99	2.3	5	NADH dehydrogenase
635	2	S22957	99	2.3	5	hypothetical prote
636	2	T39208	99	2.3	5	very hypothetical
637	2	VKLJG2	100	2.3	5	trans-regulatory s
638	1	VKLJCA	100	2.3	5	trans-regulatory s
639	2	S34400	100	2.3	5	NADH dehydrogenase
640	2	S48180	100	2.3	5	mabinlin IV - Yunn
641	2	S53097	100	2.3	5	rev protein - huma
642	2	E69460	100	2.3	5	conserved hypothet
643	2	S72816	100	2.3	5	hypothetical prote
644	2	G69268	100	2.3	5	hypothetical prote
645	2	C72655	100	2.3	5	hypothetical prote
646	4	S32195	101	2.3	5	hypothetical prote
647	1	MNVUMB	101	2.3	5	nonstructural prot
648	1	MNVUBV	101	2.3	5	nonstructural prot
649	2	B47624	101	2.3	5	Ig heavy chain V-I
650	1	G71017	101	2.3	5	hypothetical prote
651	2	S78181	102	2.3	5	NADH dehydrogenase
652	2	C41846	102	2.3	5	recombination prot
653	2	D41846	102	2.3	5	recombination prot
654	2	JE0036	102	2.3	5	hypothetical iss p
655	2	S00739	102	2.3	5	hypothetical prote
656	2	A34862	102	2.3	5	probable adhesion
657	2	T39580	102	2.3	5	Uvi31p - fission y
658	2	T16931	102	2.3	5	hypothetical prote
659	2	E72479	102	2.3	5	hypothetical prote
660	2	VKLJGG	103	2.3	5	trans-regulatory s
661	1	S12158	103	2.3	5	rev protein - huma
662	2	T05290	103	2.3	5	hypothetical prote
663	2	B34862	103	2.3	5	probable adhesion
664	2	B71053	103	2.3	5	hypothetical prote
665	2	A71071	103	2.3	5	hypothetical prote
666	2	S15395	104	2.3	5	tissue kallikrein-
667	2	A70665	104	2.3	5	probable ureB prot
668	2	S48178	104	2.3	5	mabinlin III - Yun
669	2	A42721	104	2.3	5	recombination prot
670	2	B42721	104	2.3	5	recombination prot
671	2	C42721	104	2.3	5	recombination prot
672	2	E42721	104	2.3	5	recombination prot
673	2	F42721	104	2.3	5	recombination prot
674	2	D42721	104	2.3	5	recombination prot
675	2	S43805	104	2.3	5	hypothetical prote
676	2	A75348	104	2.3	5	conserved hypothet
677	2	T41243	104	2.3	5	replication factor
678	2	T29965	104	2.3	5	hypothetical prote
679	2	H72553	104	2.3	5	hypothetical prote
680	2	S76849	105	2.3	5	hypothetical prote
681	2	F72658	105	2.3	5	hypothetical prote
682	2	H71242	106	2.3	5	hypothetical prote
683	1	VKLJST	107	2.3	5	trans-regulatory s
684	2	S23472	107	2.3	5	cpcF protein - Fis
685	2	E72710	107	2.3	5	hypothetical prote
686	1	H64913	108	2.3	5	probable membrane

No.	score	len	idx	N	Accession	Description
687	2.3	5	108		S73448	hypothetical prote
688	2.3	5	108	1	S40149	integrin alpha-7C
689	2.3	5	108		S69863	hypothetical prote
690	2.3	5	108		T06898	hypothetical prote
691	2.3	5	108		T38936	non-histone chromo
692	2.3	5	108		S65757	hypothetical prote
693	2.3	5	109		T03965	allergenic protein
694	2.3	5	109		S29655	juvenile-hormone e
695	2.3	5	109		A72622	hypothetical prote
696	2.3	5	110	1	UYBO	ubiquinone-binding
697	2.3	5	110		S22890	T-cell receptor al
698	2.3	5	110		H71472	probable sigma reg
699	2.3	5	110		T12730	hypothetical prote
700	2.3	5	110		T08921	hypothetical prote
701	2.3	5	111		A38529	nikA protein - Esc
702	2.3	5	110		D70899	hypothetical prote
703	2.3	5	111		A32450	ubiquinone-binding
704	2.3	5	111		S57889	T cell receptor Er
705	2.3	5	111		D71507	probable L23 ribos
706	2.3	5	111		C70024	nifU protein homol
707	2.3	5	111		C69168	conserved hypothet
708	2.3	5	111		B75605	hypothetical prote
709	2.3	5	111		T36016	probable electron
710	2.3	5	111		C72714	glutathione transf
711	2.3	5	112		S14355	glutathione transf
712	2.3	5	112		G70747	probable nitrogen
713	2.3	5	112		D69543	hypothetical prote
714	2.3	5	112		F75628	hypothetical prote
715	2.3	5	112	4	S62170	hypothetical prote
716	2.3	5	113		T09627	positive transcrip
717	2.3	5	113		C70691	hypothetical prote
718	2.3	5	114		S00996	Ig kappa chain pre
719	2.3	5	114		A55010	neutrophil-activat
720	2.3	5	114		F69782	hypothetical prote
721	2.3	5	115		S39393	glutathione transf
722	2.3	5	115		T28769	hypothetical prote
723	2.3	5	116		S44709	opacity protein op
724	2.3	5	116		PC4092	alanine carrier pr
725	2.3	5	116		S57143	hypothetical prote
726	2.3	5	116		T14876	hypothetical prote
727	2.3	5	116		E70510	hypothetical prote
728	2.3	5	116		B64451	hypothetical prote
729	2.3	5	116		B69270	hypothetical prote
730	2.3	5	116		T05517	abscisic acid-indu
731	2.3	5	116		D72471	hypothetical prote
732	2.3	5	117		S21668	Ig kappa chain v r
733	2.3	5	117		D71983	aspartate 1-decarb
734	2.3	5	117		B64524	aspartate 1-decarb
735	2.3	5	117		T39528	very hypothetical
736	2.3	5	117		F72669	hypothetical
737	2.3	5	117		B72508	hypothetical prote
738	2.3	5	118	1	S30566	hypothetical prote
739	2.3	5	118		G71833	beta-2-microglobul
740	2.3	5	118		T09487	ribosomal protein
741	2.3	5	118		S75220	hypothetical prote
742	2.3	5	118		E69047	hypothetical prote
743	2.3	5	118		G64682	conserved hypothet
744	2.3	5	119		S44711	apoptosis inhibito
745	2.3	5	118		H72677	opacity protein op
746	2.3	5	118		A72474	hypothetical prote
747	2.3	5	118		T14396	hypothetical prote
748	2.3	5	119		B27588	lipid transfer pro
749	2.3	5	119		A27588	Ig kappa chain pre
750	2.3	5	119		G64682	ribosomal protein
751	2.3	5	119		S54621	probable membrane
752	2.3	5	119		E75319	ferredoxin - Deino
753	2.3	5	119		T35859	hypothetical prote
754	2.3	5	119		T36326	probable dihydrone
755	2.3	5	119		T28765	hypothetical 13.7k
756	2.3	5	119		JC4546	dihydroneopterin a
757	2.3	5	119		E37854	Ig kappa chain pre
758	2.3	5	120		C29775	Ig kappa chain pre
759	2.3	5	120		E29775	Ig kappa chain pre

Description	Accession	N	idx	score	len	No.
ribosomal protein	D69696	2	120	2.3	5	760
lsu ribosomal prot	F75171	2	120	2.3	5	761
urf120 - Paracoccu	B42573	2	120	2.3	5	762
probable transcrip	A70799	2	120	2.3	5	763
protein kinase C s	S29267	2	120	2.3	5	764
hypothetical prote	T39807	3	120	2.3	5	765
carcinoembryonic a	D32268	2	121	2.3	5	766
hypothetical prote	H71086	2	121	2.3	5	767
orf6 protein - Aut	F44221	2	121	2.3	5	768
hypothetical prote	B70971	2	122	2.3	5	769
hypothetical prote	C70983	2	122	2.3	5	770
hypothetical prote	T29631	2	122	2.3	5	771
ribonuclease P (EC	D41870	2	123	2.3	5	772
probable membrane	S70044	2	123	2.3	5	773
guanine-nucleotide	I52427	2	123	2.3	5	774
guanine-nucleotide	S29714	2	123	2.3	5	775
glucagon 1 precurs	GCAF	1	124	2.3	5	776
hypothetical prote	G70785	2	124	2.3	5	777
hypothetical prote	D71100	2	124	2.3	5	778
hypothetical prote	T39259	2	124	2.3	5	779
s phase inhibitor	S71888	2	124	2.3	5	780
Spdlp - fission ye	T38491	2	124	2.3	5	781
ribosomal protein	S41139	2	125	2.3	5	782
ribosomal protein	B70307	2	125	2.3	5	783
ribosomal protein	T11941	2	125	2.3	5	784
hypothetical prote	T05147	2	125	2.3	5	785
hypothetical prote	T05193	2	125	2.3	5	786
fibroblast growth	A53627	2	125	2.3	5	787
hypothetical prote	B72521	2	125	2.3	5	788
cytochrome P450 pr	PC2257	2	126	2.3	5	789
oogenesis required	A38154	2	126	2.3	5	790
hypothetical prote	S32906	2	126	2.3	5	791
abscisic acid- and	T09789	2	126	2.3	5	792
hypothetical prote	D71239	2	126	2.3	5	793
purothionin II pre	VSBH2	1	127	2.3	5	794
hypothetical prote	T03866	2	127	2.3	5	795
hypothetical prote	S42606	2	127	2.3	5	796
hypothetical prote	A71538	2	127	2.3	5	797
probable phenylala	T06985	2	128	2.3	5	798
probable deoxycyti	C72804	2	128	2.3	5	799
hypothetical prote	A75540	2	128	2.3	5	800
hypothetical prote	E64463	2	128	2.3	5	801
conserved hypothet	F69310	2	128	2.3	5	802
hypothetical prote	D72750	2	128	2.3	5	803
Ig heavy chain pre	S03534	2	129	2.3	5	804
ribosomal protein	S34553	2	129	2.3	5	805
alpha-7A integrin	I61187	2	129	2.3	5	806
hypothetical prote	T03047	2	129	2.3	5	807
hypothetical prote	D70153	2	129	2.3	5	808
hypothetical prote	A69139	2	129	2.3	5	809
ribosomal protein	R3LV11	1	130	2.3	5	810
ribosomal protein	S77479	2	130	2.3	5	811
ribosomal protein	T07518	2	130	2.3	5	812
flagellar protein	C55546	2	130	2.3	5	813
traH protein - Ent	T08532	2	130	2.3	5	814
traH protein - Esc	S22993	2	130	2.3	5	815
hypothetical prote	F70908	2	131	2.3	5	816
hypothetical prote	C70811	2	131	2.3	5	817
hypothetical prote	T12478	2	131	2.3	5	818
ribosomal protein	R3BSS1	1	131	2.3	5	819
ribosomal protein	S23511	2	131	2.3	5	820
thionin type V pre	E29774	2	131	2.3	5	821
T-cell receptor al	S75682	2	131	2.3	5	822
hypothetical prote	H75323	2	131	2.3	5	823
hypothetical prote	T26286	2	131	2.3	5	824
spermiogenesis pro	T34088	2	131	2.3	5	825
melanin-concentrat	MTON1K	1	132	2.3	5	826
melanin-concentrat	MTON2K	1	132	2.3	5	827
melanin-concentrat	B32910	2	132	2.3	5	828
melanin-concentrat	S34653	2	132	2.3	5	829
melanin-concentrat	S34654	2	132	2.3	5	830
melanin-concentrat	A32910	2	132	2.3	5	831
T-cell receptor al	S23374	2	132	2.3	5	832

#			len	n	Accession	Description
833	2.3	5	132	2	T17596	hypothetical prote
834	2.3	5	133	1	R5EG16	ribosomal protein
835	2.3	5	133	2	A23725	ubiquinol--cytochr
836	2.3	5	133	2	S78355	H+-transporting AT
837	2.3	5	133	2	JA0087	alpha-hordothionin
838	2.3	5	133	2	S57885	T-cell receptor al
839	2.3	5	133	2	G45893	T-cell receptor al
840	2.3	5	133	2	S76285	hypothetical prote
841	2.3	5	133	2	A41298	kinesin-like prote
842	2.3	5	133	2	S47700	hypothetical 15.1K
843	2.3	5	133	2	T14238	probable secreted
844	2.3	5	133	2	PH0210	hypothetical prote
845	2.3	5	133	2	E69069	hypothetical prote
846	2.3	5	133	2	I84500	voltage-gated pota
847	2.3	5	133	2	F71189	hypothetical prote
848	2.3	5	133	2	T07357	ribosomal protein
849	2.3	5	134	2	T34591	hypothetical prote
850	2.3	5	135	2	T03288	ferredoxin [2Fe-2S
851	2.3	5	135	2	JH0684	interleukin-4 prec
852	2.3	5	135	2	JH0188	ribosomal protein
853	2.3	5	135	2	S22591	hypothetical prote
854	2.3	5	135	2	S39212	gene C2 protein -
855	2.3	5	135	2	G70519	hypothetical prote
856	2.3	5	135	2	A70317	hypothetical prote
857	2.3	5	135	2	F72688	hypothetical prote
858	2.3	5	136	2	B72105	disulfide bond oxi
859	2.3	5	136	2	S22977	beta-hordothionin
860	2.3	5	136	2	S31695	alpha-1-thionin -
861	2.3	5	136	2	G72670	probable ribosomal
862	2.3	5	136	2	A70142	ribosomal protein
863	2.3	5	136	2	S78390	ribosomal protein
864	2.3	5	136	2	S46790	ribosomal protein
865	2.3	5	136	2	S69638	ribosomal protein
866	2.3	5	136	2	T10795	acyl carrier prote
867	2.3	5	136	2	E71883	hypothetical prote
868	2.3	5	136	2	C75281	hypothetical prote
869	2.3	5	136	2	T22797	hypothetical prote
870	2.3	5	137	1	IGGP1	insulin-like growt
871	2.3	5	137	2	AYSP	acyl carrier prote
872	2.3	5	137	2	S22515	thionin precursor,
873	2.3	5	137	2	S07648	thionin precursor,
874	2.3	5	137	2	B72335	(3R)-hydroxymyrist
875	2.3	5	137	2	S38478	hypothetical prote
876	2.3	5	137	2	D38163	diglyceride kinase
877	2.3	5	137	2	T07694	hypothetical prote
878	2.3	5	137	2	G70346	spore coat polysac
879	2.3	5	138	1	R3NT11	ribosomal protein
880	2.3	5	138	1	R3PM11	ribosomal protein
881	2.3	5	138	2	RJHU2	retinoic acid-bind
882	2.3	5	138	1	QQBEB2	UL73 glycoprotein
883	2.3	5	138	2	S59086	ribosomal protein
884	2.3	5	138	2	A42495	retinoic acid-bind
885	2.3	5	138	2	B27863	Ac hypothetical pr
886	2.3	5	138	2	F75120	hypothetical prote
887	2.3	5	138	2	S67240	probable membrane
888	2.3	5	138	2	S64058	probable membrane
889	2.3	5	139	2	G69191	conserved hypothet
890	2.3	5	139	1	Q2BP87	gene 2.8 protein -
891	2.3	5	139	2	A41287	Ig heavy chain pre
892	2.3	5	139	2	S31443	glycine-rich RNA-b
893	2.3	5	139	2	S73228	ribosomal protein
894	2.3	5	139	2	I53298	cellular retinoic
895	2.3	5	139	2	T18449	pathogenicity fact
896	2.3	5	139	2	C71731	hypothetical prote
897	2.3	5	139	2	T17378	hypothetical prote
898	2.3	5	139	2	B75307	hypothetical prote
899	2.3	5	139	2	F70892	probable lpqV prot
900	2.3	5	139	2	T41526	hypothetical prote
901	2.3	5	140	1	WMV2F4	BamHI-ORF4 protein
902	2.3	5	140	2	T01170	ferredoxin [2Fe-2S
903	2.3	5	140	2	C64651	hypothetical prote
904	2.3	5	140	2	T07753	probable cysteine
905	2.3	5	140	2	T14771	hypothetical prote
906	5	2.3	141	2	H71504	ribosomal protein
907	5	2.3	141	2	JQ1553	C2 protein - Panic
908	5	2.3	141	2	T38831	hypothetical prote
909	5	2.3	141	2	S29983	lectin II - Geodia
910	5	2.3	142	2	C64071	H+-transporting AT
911	5	2.3	142	2	F71030	hypothetical prote
912	5	2.3	142	2	S39708	ywdC protein - Bac
913	5	2.3	143	1	E1HUND	Ig heavy chain pre
914	5	2.3	143	2	R3RZ11	ribosomal protein
915	5	2.3	143	2	D71911	ribonuclease hi -
916	5	2.3	143	2	E64602	ribonuclease H - H
917	5	2.3	143	2	S58584	ribosomal protein
918	5	2.3	143	2	S16248	heat shock protein
919	5	2.3	143	2	A70760	hypothetical prote
920	5	2.3	144	2	T07175	ferredoxin [2Fe-2S
921	5	2.3	144	2	A29101	vasopressin / neur
922	5	2.3	144	2	E64156	hypothetical prote
923	5	2.3	144	2	G71959	neutrophil-activat
924	5	2.3	144	2	C64550	neutrophil activat
925	5	2.3	144	2	A70500	hypothetical prote
926	5	2.3	144	2	C69527	methylmalonyl-CoA
927	5	2.3	144	2	T33109	hypothetical prote
928	5	2.3	145	2	S40831	hypothetical 15.9K
929	5	2.3	145	2	S77196	hypothetical prote
930	5	2.3	145	2	C49852	host-lethality pro
931	5	2.3	145	2	G75465	conserved hypothet
932	5	2.3	145	2	T21691	hypothetical prote
933	5	2.3	145	2	T33961	hypothetical prote
934	5	2.3	145	2	C71378	hypothetical prote
935	5	2.3	146	2	B25929	hemoglobin beta ch
936	5	2.3	146	1	ICRT4	interleukin-4 prec
937	5	2.3	147	1	QQVZCP	F7 protein - sheep
938	5	2.3	147	2	JE0148	obesity protein -
939	5	2.3	147	2	A70462	ribosomal protein
940	5	2.3	147	2	JQ1134	hypothetical prote
941	5	2.3	147	2	H75319	hypothetical prote
942	5	2.3	147	2	S15487	methylmalonyl-CoA
943	5	2.3	147	2	A69866	hypothetical prote
944	5	2.3	147	2	A69198	hypothetical prote
945	5	2.3	147	2	T04061	hypothetical prote
946	5	2.3	147	2	T09621	membrane channel p
947	5	2.3	147	2	T20615	hypothetical prote
948	5	2.3	147	2	T29241	hypothetical prote
949	5	2.3	147	2	A24580	gastrula-specific
950	5	2.3	148	2	S40148	integrin alpha-7A
951	5	2.3	148	2	S39403	hupG protein - Bra
952	5	2.3	148	2	E69151	formate hydrogenly
953	5	2.3	148	2	T39949	probable copper tr
954	5	2.3	149	2	T04171	heat shock protein
955	5	2.3	149	2	A71074	probable methylmal
956	5	2.3	149	2	S12192	hypothetical prote
957	5	2.3	149	2	D69401	hypothetical prote
958	5	2.3	149	2	S13551	stem-specific prot
959	5	2.3	150	2	T03286	ferredoxin [2Fe-2S
960	5	2.3	150	2	S20874	heat shock protein
961	5	2.3	150	2	S72545	heat shock protein
962	5	2.3	150	2	F71473	probable hydrolase
963	5	2.3	150	2	B75296	hypothetical prote
964	5	2.3	150	2	D75354	hypothetical prote
965	5	2.3	150	2	F75070	ribosomal protein
966	5	2.3	151	1	S36423	ribosomal protein
967	5	2.3	151	1	A64359	ribosomal protein
968	5	2.3	151	2	HHWT17	heat shock protein
969	5	2.3	151	2	S21600	heat shock protein
970	5	2.3	151	2	T05739	probable heat shoc
971	5	2.3	151	2	E72806	gp55 protein - Myc
972	5	2.3	151	2	T14714	reverse transcript
973	5	2.3	151	2	D70007	hypothetical prote
974	5	2.3	151	2	G70903	hypothetical prote
975	5	2.3	151	2	B70862	hypothetical prote
976	5	2.3	151	2	T38838	hypothetical prote
977	5	2.3	151	2	S37812	hypothetical prote
978	5	2.3	152	1	AWRT	atrial natriuretic

979	5	2.3	S57427	2	cysteine proteinas
980	5	2.3	S23212	2	heat shock protein
981	5	2.3	S72546	2	heat shock protein
982	5	2.3	S75230	2	hypothetical prote
983	5	2.3	D75485	2	conserved hypothet
984	5	2.3	S56198	2	hypothetical prote
985	5	2.3	T27167	2	hypothetical prote
986	5	2.3	XKPOC1	1	proteinase inhibit
987	5	2.3	HHSY17	1	heat shock protein
988	5	2.3	S23529	2	heat shock protein
989	5	2.3	T07624	2	heat shock protein
990	5	2.3	C70958	2	hypothetical prote
991	5	2.3	T09832	2	water-stress-induc
992	5	2.3	T14303	2	heat shock protein
993	5	2.3	T31701	2	hypothetical prote
994	5	2.3	R5HS30	1	ribosomal protein
995	5	2.3	JC1192	2	superoxide dismuta
996	5	2.3	A23591	2	proteinase inhibit
997	5	2.3	JS0710	2	heat shock protein
998	5	2.3	S24396	2	heat shock protein
999	5	2.3	T07629	2	small heat shock p
1000	5	2.3	S28509	2	E6 protein - Old W

ALIGNMENTS

RESULT 1
D64858
probable pseudouridylate synthase ymfC - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: D64858
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cc
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: D64858
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-207 <BLAT>
A;Cross-references: GB:AE000213; GB:U00096; NID:g1787371; PIDN:AAC74219.1; PID:g1787380;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ymfC

Query Match   23.4%;   Score 52;   DB 2;   Length 207;
Best Local Similarity   100.0%;   Pred. No. 3.1e-48;
Matches   52;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   151 WLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMG 202
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139 WLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMG 190

RESULT 2
I64156
probable pseudouridylate synthase HI0694 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C;Accession: I64156
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: I64156
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A;Residues: 1-240 <TIGR>
A;Cross-references: GB:U32752; GB:L42023; NID:g1573692; PIDN:AAC22354.1; PID:g1573697
C;Superfamily: conserved hypothetical protein HI1243

Query Match   6.8%;   Score 15;   DB 2;   Length 240;
Best Local Similarity   100.0%;   Pred. No. 3.3e-08;
Matches   15;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   74 VYAAGRLDRDSEGLL 88
             ||||||||||||||
Db   84 VYAAGRLDRDSEGLL 98

RESULT 3
S74428
probable pseudouridylate synthase slr0612 - Synechocystis sp. (PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S74428
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74428
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-185 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAA16580.1; PID:d101
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

Query Match   5.4%;   Score 12;   DB 2;   Length 185;
Best Local Similarity   100.0%;   Pred. No. 4.5e-05;
Matches   12;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   176 EGRNRQVRRMTA 187
             ||||||||||||
Db   99 EGRNRQVRRMTA 110

RESULT 4
S45555
conserved hypothetical protein ypuL - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 29-Sep-1999
C;Accession: S45555; A69943
R;Sorokin, A.; Zumstein, E.; Azevedo, V.; Ehrlich, S.D.; Serror, P.
submitted to the EMBL Data Library, November 1993
A;Reference number: S45533
A;Accession: S45555
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <SOR>
A;Cross-references: EMBL:L09228; NID:g410114; PIDN:AAA67493.1; PID:g410137
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.;
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69943
A;Accession: A69580; MUID:98044033
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-229 <KUN>
A;Cross-references: GB:299116; GB:AL009126; NID:g2634723; PIDN:CAB14248.1; PID:e1185585;
A;Experimental source: strain 168
C;Genetics:
A;Gene: ypuL
C;Superfamily: conserved hypothetical protein HI1243

Query Match   4.5%;   Score 10; DB 2; Length 229;
Best Local Similarity 100.0%;   Pred. No. 0.008;
Matches 10; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 176 EGRNRQVRRM 185
        |||||||||
Db 184 EGRNRQVRRM 193

RESULT 5
D75462
ribosomal large subunit pseudouridine synthase B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
A;Accession: D75462
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250
A;Accession: D75462
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-257 <WHI>
A;Cross-references: GB:AE001942; GB:AE000513; NID:g645861; PIDN:AAF10472.1; PID:g645861
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0896
A;Map position: 1

Query Match   4.5%;   Score 10; DB 2; Length 257;
Best Local Similarity 100.0%;   Pred. No. 0.0089;
Matches 10; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 176 EGRNRQVRRM 185
        |||||||||
Db 181 EGRNRQVRRM 190

RESULT 6
T19185
hypothetical protein C10C6.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A;Accession: T19185
R;White, S.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19086
A;Accession: T19185
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-341 <WIL>
A;Cross-references: EMBL:Z83217; PIDN:CAB05680.1; GSPDB:GN00022; CESP:C10C6.3
A;Experimental source: clone C10C6
C;Genetics:
A;Gene: CESP:C10C6.3
A;Map position: 4

A;Introns: 14/3; 81/2; 137/2; 211/2; 261/3

Query Match   3.6%;   Score 8; DB 2; Length 341;
Best Local Similarity 100.0%;   Pred. No. 1.7;
Matches 8; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 113 VQVGEGEPD 120
       |||||||
Db   45 VQVGEGEPD 52

RESULT 7
A71347
hypothetical protein TP0260 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 24-Nov-1999
A;Accession: A71347
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770
A;Accession: A71347
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-448 <COL>
A;Cross-references: GB:AE001207; GB:AE000520; NID:g3322526; PIDN:AAC65253.1; PID:g332
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0260
C;Superfamily: syphilis spirochete hypothetical protein TP0260

Query Match   3.6%;   Score 8; DB 2; Length 448;
Best Local Similarity 100.0%;   Pred. No. 2.2;
Matches 8; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   55 QFTDEAGR 62
      ||||||||
Db 287 QFTDEAGR 294

RESULT 8
B41492
58K antigen - Rickettsia tsutsugamushi
C;Species: Rickettsia tsutsugamushi
C;Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 26-Aug-1999
A;Accession: B41492
R;Stover, C.K.; Marana, D.P.; Dasch, G.A.; Oaks, E.V.
Infect. Immun. 58, 1360-1368, 1990
A;Title: Molecular cloning and sequence analysis of the sta58 major antigen gene of R
proteins.
A;Reference number: A41492; MUID:9021605
A;Accession: B41492
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-555 <STO>
A;Cross-references: GB:M31887; NID:g152501; PIDN:AAA26393.1; PID:g152503
C;Superfamily: chaperonin groEL

Query Match   3.6%;   Score 8; DB 2; Length 555;
Best Local Similarity 100.0%;   Pred. No. 2.7;
Matches 8; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 124 LAKLRNGV 131
      ||||||||
Db 372 LAKLRNGV 379

RESULT 9

B75621
sensor histidine kinase/response regulator - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: B75621
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; , M.; Shen, M.; Vamachevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250
A;Accession: B75621
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-577 <WHI>
A;Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12582.1; PID:g6460878; TIGR:DRB00
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRB0028
A;Map position: megaplasmid
A;Genome: plasmid
A;Note: plasmid MP1

Query Match   3.6%;   Score 8; DB 2; Length 577;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches   8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   58 DEAGRSTL 65
         ||||||||
Db   64 DEAGRSTL 71

RESULT 10
JE0169
dnaJ heat shock protein MCG18 - human
N;Alternate names: HSPF2
C;Species: Homo sapiens (man)
C;Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C;Accession: JE0169
R;Silins, G.; Grimmond, S.; Hayward, N.
Biochem. Biophys. Res. Commun. 243, 273-276, 1998
A;Title: Characterisation of a new human and murine member of the DnaJ family of protein
A;Reference number: JE0169; MUID:98139904
A;Accession: JE0169
A;Molecule type: mRNA
A;Residues: 1-241 <SIL>
A;Cross-references: GB:AF036874
C;Genetics:
A;Gene: MCG18
A;Map position: 11q13
A;Introns: 29/2 60/3 176/2 205/2
C;Superfamily: dnaJ amino-terminal homology
F;34-102/Domain: dnaJ amino-terminal homology <DNJ>

Query Match   3.2%;   Score 7; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 15;
Matches   7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   152 LWPRNPP 158
         |||||||
Db   12 LWPRNPP 18

RESULT 11
A43662
replicating protein repA - Agrobacterium tumefaciens plasmid pTAR
C;Species: Agrobacterium tumefaciens
C;Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999
C;Accession: A43662
R;Gallie, D.R.; Kado, C.I.
J. Bacteriol. 170, 3170-3176, 1988

A;Title: Minimal region necessary for autonomous replication of pTAR.
A;Reference number: A43662; MUID:88257036
A;Accession: A43662
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <GAL>
A;Cross-references: GB:M21299; NID:g142263; PIDN:AAD15307.1; PID:g142266
C;Genetics:
A;Genome: plasmid

Query Match   3.2%;   Score 7; DB 2; Length 250;
Best Local Similarity 100.0%; Pred. No. 15;
Matches   7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   102 QPGKRTG 108
         |||||||
Db   59 QPGKRTG 65

RESULT 12
S41382
pobR protein - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C;Accession: S41382
R;Entsch, B.; Squire, L.; Wicks, R.E.
submitted to the EMBL Data library, December 1993
A;Description: Gene for the regulation of PARA-hydroxybenzoate hydroxylase in Pseudom
A;Reference number: S41380
A;Accession: S41382
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-288 <ENT>
A;Cross-references: EMBL:X76994; NID:g444024; PIDN:CAA54301.1; PID:g444027

Query Match   3.2%;   Score 7; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 17;
Matches   7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   113 VQVEGEP 119
         |||||||
Db   63 VQVEGEP 69

RESULT 13
S74669
hypothetical protein sll1664 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S74669
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74669
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <KAN>
A;Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAA16821.1; PID:d101
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match   3.2%;   Score 7; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 19;
Matches   7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   71 VQGVYAA 77
         |||.|||

Db   298 VQGVYAA 304

RESULT 14
A71965
hypothetical protein jhp0165 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557
A;Accession: A71965
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-406 <ARN>
A;Cross-references: GB:AE001455; GB:AE001439; NID:g4154678; PIDN:AAD05746.1; PID:g415468
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0165

Query Match   3.2%; Score 7; DB 2; Length 406;
Best Local Similarity 100.0%;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   210 ANGEWRD 216
           |||||||
Db   356 ANGEWRD 362

RESULT 15
T35794
probable 4-aminobutyrate aminotransferase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 03-Dec-1999
C;Accession: T35794
R;Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A;Reference number: 221556
A;Accession: T35794
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-444 <MUR>
A;Cross-references: EMBL:AL031225; PIDN:CAA20213.1; GSPDB:GN00070; SCOEDB:SC8B7.02
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: gabT; SCOEDB:SC8B7.02

Query Match   3.2%; Score 7; DB 2; Length 444;
Best Local Similarity 100.0%;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   85 EGLLVLT 91
           |||||||
Db   403 EGLLVLT 409

Search completed: June 10, 2000, 11:27:08
Job time: 2521 sec

THIS PAGE BLANK (LEFT)

OM protein - protein search, using sw model

```
                  GenCore version 4.5
                  Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:           June 10, 2000, 11:27:09 ; Search time 9.96 Seconds
                                             (without alignments)
                                             678.816 Million cell updates/sec

Title:            US-09-252-691-7056
Perfect score:    222
Sequence:         1 AIMRQLITPENTMKTSFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table:    OLIGO
                  Gapop 60.0 , Gapext 60.0
Searched:         83857 seqs, 3045973 residues
Word size :       0

Total number of hits satisfying chosen parameters:    83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database :        SwissProt_38:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52	23.4	217	1	YMFC_ECOLI	P75966 escherichia
2	15	6.8	240	1	YMFC_HAEIN	P44827 haemophilus
3	12	5.4	261	1	Y612_SYNY3	P72581 synechocyst
4	10	4.5	229	1	RLUB_BACSU	P35159 bacillus su
5	8	3.6	555	1	CH60_RICTS	P16625 rickettsia
6	8	3.2	250	1	REPA_AGRTU	P15394 agrobacteri
7	7	3.2	387	1	STCV_EMENI	Q00727 emericella
8	7	3.2	454	1	NFM_PIG	P08552 sus scrofa
9	7	3.2	515	1	PDI_ASPNG	Q12730 aspergillus
10	7	3.2	534	1	TCPG_YEAST	P39077 saccharomyc
11	7	3.2	546	1	TCPZ_YEAST	P39079 saccharomyc
12	7	3.2	604	1	SP20_YEAST	P50805 saccharomyc
13	7	3.2	707	1	COG9_RABIT	P41246 oryctolagus
14	7	3.2	764	1	ACOC_CUCMC	Q42669 cucumis mel
15	7	3.2	845	1	NFM_RAT	P12839 rattus norv
16	7	3.2	848	1	NFM_MOUSE	P08553 mus musculu
17	7	3.2	868	1	ACOC_CUCMA	P49608 cucurbita m
18	7	3.2	915	1	NFM_HUMAN	P07197 homo sapien
19	7	3.2	990	1	TRA7_ECOLI	P13694 escherichia
20	7	3.2	1148	1	MFD_ECOLI	P30958 escherichia
21	7	3.2	1886	1	GP21_RAT	P11654 rattus norv
22	6	2.7	89	1	VFUS_ORFN2	P26654 orf virus (
23	6	2.7	94	1	IHFB_PSEAE	Q51473 pseudomonas
24	6	2.7	97	1	Y109_EMENI	P05678 emericella
25	6	2.7	103	1	CEAV_ECOLI	P22522 escherichia
26	6	2.7	106	1	YNIU_RHOCA	Q07184 rhodobacter
27	6	2.7	133	1	IL4_PIG	Q04745 sus scrofa
28	6	2.7	134	1	H1_EUPEU	P32103 euplotes eu
29	6	2.7	142	1	YFCC_CHRVI	Q06536 chromatium
30	6	2.7	150	1	FKBX_PSEFL	P21863 pseudomonas
31	6	2.7	158	1	UREE_KLEAE	P18317 klebsiella
32	6	2.7	165	1	LKTC_PASSP	P55124 pasteurella
33	6	2.7	173	1	CRGD_HUMAN	P07320 homo sapien
34	6	2.7	174	1	PLU_DROME	P42570 drosophila
35	6	2.7	191	1	APL2_PETMA	P07096 petromyzon
36	6	2.7	192	1	BGLS_SCHCO	P29091 schizophyll
37	6	2.7	193	1	YHHF_HAEIN	P44869 haemophilus
38	6	2.7	204	1	CRB2_BOVIN	P02522 bos taurus
39	6	2.7	207	1	FRDA_MOUSE	O35943 mus musculu
40	6	2.7	209	1	LOLB_HAEIN	P45270 haemophilus
41	6	2.7	210	1	FRDA_HUMAN	Q16595 homo sapien
42	6	2.7	217	1	HXB7_HUMAN	P09629 homo sapien
43	6	2.7	217	1	HXB7_MOUSE	P09024 mus musculu
44	6	2.7	231	1	RSUA_ECOLI	P33918 escherichia
45	6	2.7	235	1	NHAB_RHOSO	Q53117 rhodococcus
46	6	2.7	238	1	HIS4_METVA	P05324 methanococc
47	6	2.7	239	1	TRPA_THEMA	P50908 thermotoga
48	6	2.7	242	1	HIS4_METVO	P05325 methanococc
49	6	2.7	247	1	TRUA_BACSU	P70973 bacillus su
50	6	2.7	249	1	Y361_SYNY3	Q55578 synechocyst
51	6	2.7	250	1	YQEE_BACSU	P54450 bacillus su
52	6	2.7	254	1	PMM_YEAST	P07283 saccharomyc
53	6	2.7	260	1	ARGT_ECOLI	P09551 escherichia
54	6	2.7	261	1	YB15_METJA	Q58515 methanococc
55	6	2.7	262	1	ATP6_ALLAR	P50363 allomyces a
56	6	2.7	262	1	ATP6_ALLMA	P50364 allomyces m
57	6	2.7	265	1	IHA_SHEEP	P38440 ovis aries
58	6	2.7	266	1	BLO7_ECOLI	P35695 escherichia
59	6	2.7	266	1	BLOB_PSEAE	Q06778 pseudomonas
60	6	2.7	266	1	BLP2_PSEAE	P14489 pseudomonas
61	6	2.7	268	1	INHA_MYCAV	O07400 mycobacteri
62	6	2.7	272	1	NADE_BACSU	P08164 bacillus su
63	6	2.7	274	1	YE03_MYCTU	P71671 mycobacteri
64	6	2.7	284	1	YNB9_YEAST	P63975 saccharomyc
65	6	2.7	284	1	YND3_YEAST	P53964 saccharomyc
66	6	2.7	288	1	SCRK_PEDPE	P43468 pediococcus
67	6	2.7	290	1	SUHB_MYCTU	O07203 mycobacteri
68	6	2.7	294	1	ARGB_CORGL	Q59281 corynebacte
69	6	2.7	304	1	PHEA_AMYME	Q44104 amycolatops
70	6	2.7	313	1	ADT_NEUCR	P02723 neurospora
71	6	2.7	313	1	YTL2_SALTY	P37415 salmonella
72	6	2.7	318	1	Y940_METJA	Q58350 methanococc
73	6	2.7	319	1	YIDZ_ECOLI	P31463 escherichia
74	6	2.7	323	1	CRO_RANCA	P17264 rana catesb
75	6	2.7	323	1	CRO_RANTE	P02532 rana tempor
76	6	2.7	327	1	DHB1_HUMAN	P14061 homo sapien
77	6	2.7	329	1	YA2E_SCHPO	Q09704 schizosacch
78	6	2.7	330	1	YHA9_YEAST	P38749 saccharomyc
79	6	2.7	339	1	PURK_BRUME	P52559 brucella me
80	6	2.7	340	1	YD60_MYCTU	Q11030 mycobacteri
81	6	2.7	346	1	CCHL_NEUCR	P14187 neurospora
82	6	2.7	353	1	VP11_RDV	P16594 rice dwarf
83	6	2.7	354	1	RECF_SALTY	P24900 salmonella
84	6	2.7	356	1	RECF_ECOLI	P03016 escherichia
85	6	2.7	356	1	SMR1_PODAN	Q08142 podospora a
86	6	2.7	359	1	RECF_HAEIN	P43767 haemophilus
87	6	2.7	359	1	YKA7_CAEEL	P34258 caenorhabdi
88	6	2.7	360	1	IHA_BOVIN	P07994 bos taurus
89	6	2.7	360	1	YA63_MYCTU	O53411 mycobacteri
90	6	2.7	362	1	RECF_PROMI	P22839 proteus mir
91	6	2.7	368	1	PEPQ_LACDL	P46545 lactobacill
92	6	2.7	371	1	MMK2_MEDSA	Q40353 medicago sa
93	6	2.7	377	1	YD67_MYCTU	Q11037 mycobacteri
94	6	2.7	383	1	APL_HUMAN	O14791 homo sapien
95	6	2.7	388	1	YA25_METJA	Q58431 methanococc
96	6	2.7	388	1	YC67_MYCTU	Q11052 mycobacteri
97	6	2.7	397	1	ISP7_SCHPO	P40902 schizosacch
98	6	2.7	398	1	Y106_MYCTU	Q10899 mycobacteri
99	6	2.7	399	1	RRPP_PI4HA	P22044 human parai
100	6	2.7	405	1	PAG1_TRYBB	Q01889 trypanosoma
101	6	2.7	407	1	YII6_MYCTU	Q72052 mycobacteri
102	6	2.7	413	1	MAAL_CLOTT	Q05514 clostridium
103	6	2.7	414	1	GLYA_CAMJE	Q24531 campylobact
104	6	2.7	415	1	PGK_DROME	Q01604 drosophila
105	6	2.7	420	1	YFJD_HAEIN	Q57017 haemophilus
106	6	2.7	426	1	SYY_MYCLE	Q49900 mycobacteri

Left block

#			Len		Name	Organism	Accession
107	2.7	6	428	1	CG2B_ARATH	arabidopsis	P30183
108	2.7	6	432	1	FIBX_MOUSE	mus musculu	P12804
109	2.7	6	450	1	YAEL_ECOLI	escherichia	P37764
110	2.7	6	451	1	ARGD_ALNGL	alnus gluti	O04866
111	2.7	6	452	1	YP59_MYCTU	mycobacteri	Q50739
112	2.7	6	465	1	DCDR_XENLA	xenopus lae	P42291
113	2.7	6	466	1	B1AR_MOUSE	mus musculu	P34971
114	2.7	6	466	1	B1AR_RAT	rattus norv	P18090
115	2.7	6	467	1	PPAR_CAVPO	cavia porce	O35507
116	2.7	6	474	1	DLD2_BACSU	bacillus su	P54533
117	2.7	6	488	1	DHAL_BACST	bacillus st	P42329
118	2.7	6	499	1	VL1_HPV13	human papil	Q02273
119	2.7	6	499	1	VL1_HPV33	human papil	P06416
120	2.7	6	499	1	VL1_HPV53	human papil	Q05113
121	2.7	6	500	1	CPJ1_RABIT	oryctolagus	P52786
122	2.7	6	500	1	VL1_HPV6A	human papil	P95787
123	2.7	6	501	1	ATPA_STRMU	streptococc	P04012
124	2.7	6	501	1	VL1_HPV11	human papil	P22163
125	2.7	6	501	1	VL1_RHPV1	rhesus papi	P50791
126	2.7	6	502	1	VL1_HPV28	human papil	P27233
127	2.7	6	502	1	VL1_HPV42	human papil	Q02274
128	2.7	6	502	1	VL1_PCPV1	pygmy chimp	P36735
129	2.7	6	503	1	VL1_HPV26	human papil	Q80961
130	2.7	6	503	1	VL1_HPV66	human papil	P50793
131	2.7	6	504	1	VL1_HPV70	human papil	P24838
132	2.7	6	505	1	VL1_HPV39	human papil	P50822
133	2.7	6	505	1	VL1_HPV61	human papil	P54669
134	2.7	6	505	1	VL1_HPV68	human papil	P27964
135	2.7	6	505	1	VL1_HPVME	human papil	Q02515
136	2.7	6	508	1	VL1_HPV30	human papil	Q25486
137	2.7	6	510	1	VL1_HPV2A	human papil	P22162
138	2.7	6	510	1	VL1_HPV57	mus musculu	Q60755
139	2.7	6	515	1	CALR_MOUSE	saccharomyc	P15202
140	2.7	6	515	1	CATA_YEAST	enterococcu	P26679
141	2.7	6	518	1	ATPA_ENTHR	thiocapsa r	P52674
142	2.7	6	522	1	CYSJ_THIRO	human papil	P26535
143	2.7	6	524	1	VL1_HPV58	human papil	Q05138
144	2.7	6	529	1	VL1_HPV52	human papil	P36732
145	2.7	6	531	1	VL1_HPV10	human papil	P36731
146	2.7	6	532	1	VL1_HPV03	human papil	P36743
147	2.7	6	539	1	VL1_HPV56	human papil	P36741
148	2.7	6	539	1	VL1_HPV45	human papil	P14913
149	2.7	6	544	1	4CL2_PETCR	petroselinu	Q01770
150	2.7	6	544	1	PRIS_DESDE	desulfovibr	O29248
151	2.7	6	547	1	ILVD_ARCFU	archaeoglob	Q49375
152	2.7	6	548	1	ATPA_KULA	kluyveromyc	Q58175
153	2.7	6	548	1	PRIS_METJA	methanococc	P31101
154	2.7	6	553	1	PRIS_DESVH	desulfovibr	P00962
155	2.7	6	553	1	SYQ_ECOLI	escherichia	P52388
156	2.7	6	554	1	UL25_HSV7J	herpes simp	Q10351
157	2.7	6	554	1	YDA9_SCHPO	schizosacch	O22212
158	2.7	6	554	1	YG62_ARATH	arabidopsis	P96095
159	2.7	6	556	1	PRIS_THIFE	thiobacillu	O00782
160	2.7	6	559	1	TCPG_OXYGR	oxytricha g	P06794
161	2.7	6	568	1	VL1_HPV18	human papil	P06115
162	2.7	6	573	1	CATT_YEAST	saccharomyc	P39693
163	2.7	6	575	1	RECJ_ERWCH	erwinia chr	P45112
164	2.7	6	575	1	RECJ_HAEIN	haemophilus	P21893
165	2.7	6	577	1	RECJ_ECOLI	escherichia	P36736
166	2.7	6	584	1	VL1_HPV27	human papil	Q9zdq1
167	2.7	6	600	1	LEPA_RICPR	rickettsia	P37295
168	2.7	6	602	1	PSY_NEUCR	neurospora	P93107
169	2.7	6	606	1	PF20_CHLRE	chlamydomon	P19721
170	2.7	6	613	1	FIBP_ADEM1	mouse adeno	P48065
171	2.7	6	614	1	NTBE_HUMAN	homo sapien	Q60696
172	2.7	6	626	1	NTBE_RABIT	oryctolagus	Q60714
173	2.7	6	646	1	PM17_MOUSE	mus musculu	P15379
174	2.7	6	655	1	FATP_MOUSE	mus musculu	O83258
175	2.7	6	657	1	CD44_MOUSE	treponema p	P53165
176	2.7	6	657	1	PRIA_TREPA	saccharomyc	P40967
177	2.7	6	661	1	YGG6_YEAST	homo sapien	P10505
178	2.7	6	661	1	PM17_HUMAN	homo sapien	
179	2.7	6	665	1	NUC2_SCHPO	schizosacch	

Right block

#			Len		Name	Accession	Organism
180	6	2.7	669	1	GSH1_SCHPO	Q09768	schizosacch
181	6	2.7	677	1	SG1_MOUSE	P16014	mus musculu
182	6	2.7	677	1	YPK2_YEAST	P18961	saccharomyc
183	6	2.7	709	1	SYR_CAEEL	Q19825	caenorhabdi
184	6	2.7	722	1	MESD_LEUME	Q10418	leuconostoc
185	6	2.7	746	1	GUN1_STRRE	Q05156	streptomyce
186	6	2.7	775	1	VP4_ROTH3	P39033	human rotav
187	6	2.7	775	1	VP4_ROTH5	P30214	human rotav
188	6	2.7	779	1	GEF1_YEAST	P37020	saccharomyc
189	6	2.7	810	1	MECB_BACSU	P37571	bacillus su
190	6	2.7	823	1	YHX2_YEAST	P38863	saccharomyc
191	6	2.7	847	1	HEX_VIBVU	Q04786	vibrio vuln
192	6	2.7	856	1	PPSA_AQUAE	Q67899	aquifex aeo
193	6	2.7	859	1	LUXO_VIBHA	P54302	vibrio harv
194	6	2.7	862	1	LOX1_HORVU	P29114	hordeum vul
195	6	2.7	863	1	APCE_CYACA	P35911	cyanidium c
196	6	2.7	865	1	LOX2_ORYSA	P29250	oryza sativ
197	6	2.7	869	1	PEM1_YEAST	P05374	saccharomyc
198	6	2.7	878	1	IL3B_MOUSE	P26954	mus musculu
199	6	2.7	879	1	FPRP_RAT	Q62786	rattus norv
200	6	2.7	896	1	CYRB_MOUSE	P26955	mus musculu
201	6	2.7	897	1	CYRB_HUMAN	P32927	homo sapien
202	6	2.7	909	1	UBPB_SCHPO	Q09879	schizosacch
203	6	2.7	910	1	HUL5_YEAST	P53119	saccharomyc
204	6	2.7	910	1	YK69_YEAST	P36165	saccharomyc
205	6	2.7	914	1	GUX2_CLOSR	P50900	clostridium
206	6	2.7	923	1	AGLU_TETPY	O00906	tetrahymena
207	6	2.7	946	1	RGC1_HUMAN	P98171	homo sapien
208	6	2.7	947	1	PODK_MAIZE	P11155	zea mays (m
209	6	2.7	949	1	PODK_MESCR	Q42910	mesembryant
210	6	2.7	953	1	PODK_FLABI	Q39735	flaveria bi
211	6	2.7	953	1	PODK_FLATR	P22221	flaveria tr
212	6	2.7	955	1	PODK_FLABR	Q39734	flaveria br
213	6	2.7	956	1	PODK_FLAPR	Q42736	flaveria pr
214	6	2.7	971	1	VP2_EHDV1	Q00273	epizootic h
215	6	2.7	979	1	VGLB_HSVEL	P25218	equine herp
216	6	2.7	980	1	VGLB_HSVE1	P18050	equine herp
217	6	2.7	980	1	VGLB_HSVEA	P18551	equine herp
218	6	2.7	980	1	VGLB_HSVEB	P28922	equine herp
219	6	2.7	1008	1	GI10_YEAST	P36048	saccharomyc
220	6	2.7	1033	1	YDK9_SCHPO	P87115	schizosacch
221	6	2.7	1039	1	SYI_METJA	Q58357	methanococc
222	6	2.7	1081	1	KOKO_YEAST	Q12236	saccharomyc
223	6	2.7	1123	1	PHYA_SOLTU	P30733	solanum tub
224	6	2.7	1124	1	PHY1_TOBAC	P33530	nicotiana t
225	6	2.7	1157	1	RIR1_HSV23	P09853	herpes simp
226	6	2.7	1174	1	Y182_HUMAN	Q14687	homo sapien
227	6	2.7	1179	1	KCRF_STRPU	P18294	strongyloce
228	6	2.7	1227	1	RPOB_MYCLE	P30760	mycobacteri
229	6	2.7	1239	1	LAF4_HUMAN	P51826	homo sapien
230	6	2.7	1242	1	CHS5_USTMA	O13394	ustilago ma
231	6	2.7	1290	1	IRS1_HUMAN	P35568	homo sapien
232	6	2.7	1290	1	PIP4_HUMAN	P19174	homo sapien
233	6	2.7	1291	1	PIP4_RAT	P10686	rattus norv
234	6	2.7	1330	1	VCAP_PRVIS	P08487	bos taurus
235	6	2.7	1344	1	IF3A_MOUSE	Q00705	pseudorable
236	6	2.7	1394	1	IF3A_HUMAN	P23116	mus musculu
237	6	2.7	1394	1	CNG4_BOVIN	Q14152	homo sapien
238	6	2.7	1475	1	TRA2_CAEEL	Q28181	bos taurus
239	6	2.7	1481	1	APU_THEET	P34709	caenorhabdi
240	6	2.7	1497	1	TRA2_CAEBR	P38939	t amylopull
241	6	2.7	1515	1	GDE_HUMAN	Q17307	caenorhabdi
242	6	2.7	1555	1	YM81_YEAST	P35573	homo sapien
243	6	2.7	1562	1	ICEN_XANCT	P35574	saccharomyc
244	6	2.7	1567	1	FAS2_SCHPO	Q04701	xanthomonas
245	6	2.7	1842	1	RPB1_CAEEL	P18127	schizosacch
246	6	2.7	1859	1	BIME_EMENI	Q10289	caenorhabdi
247	6	2.7	2073	1	POLG_POL2W	P16356	emericella
248	6	2.7	2205	1	POLG_CXA21	P24686	poliovirus
249	6	2.7	2206	1	POLG_POL1M	P23069	coxsackievi
250	6	2.7	2206	1	POLG_POL32	P22005	p genome po
251	6	2.7	2206	1		P03299	poliovirus
252	6	2.7	2206	1		P06209	poliovirus

#			Name		Len	Organism	Accession
253	6	2.7	POLG_POL3L	1	2206	poliovirus	P03302
254	6	2.7	POLG_POL2L	1	2207	poliovirus	P06210
255	6	2.7	POLH_POLIM	1	2208	p genome po	P03300
256	6	2.7	POLG_POLIS	1	2209	poliovirus	P03301
257	6	2.7	POLG_CXA24	1	2214	coxsackievi	P36290
258	6	2.7	RRPL_MUMPM	1	2261	mumps virus	P30929
259	6	2.7	YAQ5_SCHPO	1	2670	schizosacch	Q10105
260	6	2.7	POLG_HCVH	1	3011	h genome po	P27958
261	6	2.7	IRA1_YEAST	1	3092	saccharomyc	P18963
262	6	2.7	HD_RAT	1	3110	rattus norv	P51111
263	6	2.7	HD_HUMAN	1	3144	homo sapien	P42858
264	6	2.7	POLG_TBEVH	1	3414	t genome po	Q01299
265	6	2.7	ERY2_SACER	1	3567	saccharopol	Q03132
266	6	2.7	LRP1_CHICK	1	4543	gallus gall	P98157
267	5	2.3	H1L5_ENSMI	1	36	ensis minor	P27203
268	5	2.3	GBG7_MOUSE	1	42	mus musculu	Q61016
269	5	2.3	BAGE_HUMAN	1	43	homo sapien	Q13072
270	5	2.3	DPSD_AZOVI	1	48	azotobacter	Q44558
271	5	2.3	THPO_HUMAN	1	48	homo sapien	P08918
272	5	2.3	THPS_HUMAN	1	48	homo sapien	P08919
273	5	2.3	THP1_BOVIN	1	49	bos taurus	P01249
274	5	2.3	THP2_BOVIN	1	49	bos taurus	P01250
275	5	2.3	THPS_BOVIN	1	49	bos taurus	P01251
276	5	2.3	Y195_BPT3	1	49	bacteriopha	P10306
277	5	2.3	RUBR_CHLLT	1	53	chlorobium	P09947
278	5	2.3	RL15_LACLA	1	55	lactococcus	P27145
279	5	2.3	HBX4_ECHGR	1	60	echinococcu	P55813
280	5	2.3	SECE_THETH	1	60	thermus aqu	P38383
281	5	2.3	DNBI_BFDV	1	61	budgerigar	P13893
282	5	2.3	YM26_MARPO	1	63	marchantia	P38468
283	5	2.3	GBG7_BOVIN	1	68	bos taurus	P30671
284	5	2.3	GBG7_HUMAN	1	68	homo sapien	O60262
285	5	2.3	DHAG_PIG	1	69	sus scrofa	Q29228
286	5	2.3	GBG7_RAT	1	69	rattus norv	P43425
287	5	2.3	Y375_TREPA	1	69	treponema p	O83390
288	5	2.3	FPG_RHIME	1	70	rhizobium m	Q59752
289	5	2.3	RK31_PORPU	1	70	porphyra pu	P51290
290	5	2.3	RL31_HAEDU	1	70	haemophilus	Q59450
291	5	2.3	RL31_HAEIN	1	70	haemophilus	P44367
292	5	2.3	Y4JM_RHISN	1	70	rhizobium s	P55513
293	5	2.3	RK31_GUITH	1	72	guillardia	O46917
294	5	2.3	RK31_ODOSI	1	72	odontella s	P49563
295	5	2.3	RL29_SYNY3	1	73	synechocyst	P73312
296	5	2.3	HS12_SOYBN	1	74	glycine max	P02520
297	5	2.3	RS18_NEIGO	1	76	neisseria g	O07815
298	5	2.3	RL18_ECOLI	1	76	escherichia	P02038
299	5	2.3	VG43_BPML5	1	77	mycobacteri	Q05255
300	5	2.3	Y4DJ_RHISN	1	77	rhizobium s	P55409
301	5	2.3	NU4M_SIMVI	1	79	simulium vi	P50660
302	5	2.3	COXM_HUMAN	1	80	homo sapien	P24311
303	5	2.3	COXM_MOUSE	1	80	mus musculu	P56393
304	5	2.3	ICTA_SYNY3	1	80	synechocyst	P27372
305	5	2.3	ITF_HUMAN	1	80	homo sapien	Q07654
306	5	2.3	UL11_VZVD	1	81	varicella-z	P09297
307	5	2.3	CHPR_ECOLI	1	82	escherichia	P18534
308	5	2.3	YK76_MYCTU	1	83	mycobacteri	Q10684
309	5	2.3	RL17_ODOSI	1	84	odontella s	P49504
310	5	2.3	YOEB_ECOLI	1	84	escherichia	P56605
311	5	2.3	ECTP_TRIGR	1	87	tripneustes	P23227
312	5	2.3	COXM_BOVIN	1	88	bos taurus	P13183
313	5	2.3	GRLA_LEGPN	1	89	legionella	Q48833
314	5	2.3	RS15_TREPA	1	89	treponema p	O83857
315	5	2.3	SY05_MOUSE	1	91	mus musculu	P30882
316	5	2.3	SY05_RAT	1	92	rattus norv	P50231
317	5	2.3	EM2_WHEAT	1	94	triticum ae	P22701
318	5	2.3	FE45_BRAJA	1	94	bradyrhizob	Q45218
319	5	2.3	IHFB_HAEIN	1	94	haemophilus	P43724
320	5	2.3	YOY3_CAEEL	1	94	caenorhabdi	P34661
321	5	2.3	CILG_HAEIN	1	95	haemophilus	P44461
322	5	2.3	APC3_PIG	1	96	sus scrofa	P27917
323	5	2.3	GRO_RAT	1	96	rattus norv	P14095
324	5	2.3	RR6_GUITH	1	96	guillardia	O78447
325	5	2.3	YFRC_PROVU	1	96	proteus vul	P20927

#			Score		Name	Organism	Accession
326	5	2.3	97	1	EOTA_MOUSE	mus musculu	P48298
327	5	2.3	97	1	EOTA_RAT	rattus norv	P97545
328	5	2.3	97	1	FER_ARCLA	arctium lap	P00223
329	5	2.3	97	1	KRFA_CHICK	gallus gall	P20308
330	5	2.3	97	1	MXII_SHIFL	shigella fl	Q06080
331	5	2.3	99	1	Y11K_STRFR	streptomyce	P26800
332	5	2.3	99	1	Y141_METJA	methanococc	Q57606
333	5	2.3	99	1	2SS4_CAPMA	capparis ma	P80353
334	5	2.3	100	1	REV_HV2CA	human immun	P24104
335	5	2.3	100	1	REV_HV2RO	human immun	P04615
336	5	2.3	101	1	VNST_BUNYW	bunyamwera	P16494
337	5	2.3	101	1	VNST_MAGV	maguari vir	P16605
338	5	2.3	101	1	B1_BBV	black beetl	P29099
339	5	2.3	102	1	G24A_DICDI	dictyosteli	P16642
340	5	2.3	102	1	ISS_ECOLI	escherichia	P19592
341	5	2.3	102	1	PIGR_PIG	sus scrofa	Q29244
342	5	2.3	102	1	RECA_ENTFA	enterococcu	P42444
343	5	2.3	102	1	RR10_GUITH	guillardia	P19460
344	5	2.3	102	1	TCPD_SCHPO	schizosacch	P50999
345	5	2.3	102	1	UV31_SCHPO	schizosacch	Q12238
346	5	2.3	102	1	YNI2_METTL	methanococc	P05410
347	5	2.3	103	1	G24B_DICDI	dictyosteli	P16643
348	5	2.3	103	1	REV_HV2D1	human immun	P17754
349	5	2.3	103	1	REV_HV2G1	human immun	P18039
350	5	2.3	104	1	2SS3_CAPMA	capparis ma	P80352
351	5	2.3	104	1	HKL5_ARATH	arabidopsis	P48002
352	5	2.3	104	1	RECA_CLOAB	clostridium	P42441
353	5	2.3	104	1	RECA_LACDE	lactobacill	Q02347
354	5	2.3	104	1	RECA_LACHE	lactobacill	Q02348
355	5	2.3	104	1	RECA_LEUME	leuconostoc	Q02349
356	5	2.3	104	1	RFA3_SCHPO	schizosacch	Q92374
357	5	2.3	104	1	URE2_MYCTU	mycobacteri	P50048
358	5	2.3	106	1	GAC_YEAST	saccharomyc	P33440
359	5	2.3	107	1	CPCF_MASLA	mastigoclad	P29734
360	5	2.3	107	1	REV_HV2NZ	human immun	P05873
361	5	2.3	107	1	REV_HV2ST	human immun	P20870
362	5	2.3	108	1	Y029_MYCPN	mycoplasma	P75082
363	5	2.3	108	1	YC50_CYAPA	cyanophora	P48326
364	5	2.3	108	1	YNFA_ECOLI	escherichia	P76169
365	5	2.3	109	1	JHEA_TRINI	trichoplusi	P30809
366	5	2.3	109	1	VNST_BUNGE	bunyavirus	P16992
367	5	2.3	110	1	UCR6_BOVIN	bos taurus	P00129
368	5	2.3	110	1	UCR6_HUMAN	homo sapien	P14927
369	5	2.3	110	1	YD90_MYCTU	mycobacteri	P71660
370	5	2.3	112	1	GLNB_MYCTU	mycobacteri	Q10960
371	5	2.3	113	1	RAMA_ENTCL	enterobacte	P55922
372	5	2.3	113	1	RAMA_KLEPN	klebsiella	Q48413
373	5	2.3	113	1	TRPF_BACME	bacillus me	P70938
374	5	2.3	114	1	EN78_HUMAN	homo sapien	P42830
375	5	2.3	116	1	YJ90_YEAST	saccharomyc	P47157
376	5	2.3	117	1	CHH3_PENJP	penaeus jap	Q94676
377	5	2.3	117	1	PAND_HELPY	helicobacte	P56065
378	5	2.3	117	1	RS25_CAEEL	caenorhabdi	P52821
379	5	2.3	118	1	B2MG_HORSE	equus cabal	P30441
380	5	2.3	118	1	IAP4_NPVOP	orgyia pseu	O10345
381	5	2.3	118	1	RL18_HELPJ	helicobacte	Q9zjs7
382	5	2.3	119	1	RL18_STRCO	streptomyce	Q9x810
383	5	2.3	119	1	RL18_HELPY	helicobacte	P56043
384	5	2.3	120	1	FOLB_BACSU	bacillus su	P28823
385	5	2.3	120	1	ML1B_RAT	rattus norv	P49287
386	5	2.3	120	1	RL18_BACSU	bacillus su	P46899
387	5	2.3	120	1	Y06I_BPT4	bacteriopha	P13316
388	5	2.3	121	1	NUOA_RHIME	rhizobium m	O68852
389	5	2.3	121	1	RECA_RUMAL	ruminococcu	P49985
390	5	2.3	122	1	RL17_CHLTR	chlamydia t	P47760
391	5	2.3	123	1	GLN2_METBA	methanosarc	P54809
392	5	2.3	123	1	MSS4_HUMAN	homo sapien	P47224
393	5	2.3	123	1	MSS4_RAT	rattus norv	Q08326
394	5	2.3	123	1	RNPA_STRCO	streptomyce	P48206
395	5	2.3	124	1	GLU1_LOPAM	lophius ame	P01278
396	5	2.3	124	1	YM06_MYCTU	mycobacteri	Q10395
397	5	2.3	125	1	RS11_AQUAE	aquifex aeo	O66485
398	5	2.3	125	1	RT11_MARPO	marchantia	P26870

Left column block:

No.		Len			Accession	Organism	ID
399	1	125	2.3	5	P46744	prototheca	RT12_PROWI
400	1	126	2.3	5	P25843	drosophila	PROF_DROME
401	1	127	2.3	5	P01545	hordeum vul	THNA_HORVU
402	1	127	2.3	5	P42880	rhizobium m	UREY_RHIME
403	1	128	2.3	5	O22000	mycobacteri	DCTD_BPMD2
404	1	129	2.3	5	Q45791	bacteroides	RECA_BACTN
405	1	129	2.3	5	P27419	euglena gra	RR11_EUGGR
406	1	129	2.3	5	P34403	caenorhabdi	YLU9_CAEEL
407	1	130	2.3	5	P40728	salmonella	FLHE_SALTY
408	1	130	2.3	5	P17647	cavia porce	IGF1_CAVPO
409	1	130	2.3	5	O78409	guillardia	RK19_GUITH
410	1	130	2.3	5	P06364	marchantia	RR11_MARPO
411	1	130	2.3	5	P41630	pinus thunb	RR11_PINTH
412	1	130	2.3	5	O24709	synechococc	RS11_SYNP6
413	1	130	2.3	5	P73298	synechocyst	RS11_STNY3
414	1	130	2.3	5	Q00189	synechocyst	TRH5_ECOLI
415	1	131	2.3	5	P04969	bacillus su	RS11_BACSU
416	1	131	2.3	5	P54218	caenorhabdi	SP27_CAEEL
417	1	131	2.3	5	Q05806	triticum ae	THN5_WHEAT
418	1	132	2.3	5	P19713	oncorhynchu	MLC1_ONCKE
419	1	132	2.3	5	P33745	oncorhynchu	MLC1_ONCMY
420	1	132	2.3	5	P17640	oncorhynchu	MLC1_ONCTS
421	1	132	2.3	5	P19714	oncorhynchu	MLC2_ONCKE
422	1	133	2.3	5	P49648	odontella s	ATPE_ODOSI
423	1	133	2.3	5	P55030	felis silve	IL4_FELCA
424	1	133	2.3	5	Q55625	synechocyst	RBFA_SYNY3
425	1	133	2.3	5	O33752	synechococc	YH35_SYNP2
426	1	133	2.3	5	P28910	escherichia	YHHG_ECOLI
427	1	134	2.3	5	Q00247	plectonema	YUF1_PLEBO
428	1	134	2.3	5	O50467	neisseria g	AROK_NEIGO
429	1	134	2.3	5	P56364	chlorella v	RK16_CHLVU
430	1	134	2.3	5	O46901	guillardia	RK16_GUITH
431	1	135	2.3	5	P27789	zea mays (m	FER5_MAIZE
432	1	135	2.3	5	P30367	bos taurus	IL4_BOVIN
433	1	135	2.3	5	P21512	euglena gra	RK16_EUGGR
434	1	135	2.3	5	P16633	gracilaria	VAL2_TYLCM
435	1	135	2.3	5	P27263	tomato yell	VAL2_TYLCU
436	1	135	2.3	5	P38610	tomato yell	FUR_LEGPN
437	1	136	2.3	5	Q48835	legionella	RL27_YEAST
438	1	136	2.3	5	P38706	saccharomyc	RR11_EPIVI
439	1	136	2.3	5	P30059	epifagus vi	RS8E_AERPE
440	1	136	2.3	5	Q9ydy0	aeropyrum p	RS9_BORBU
441	1	136	2.3	5	O51313	borrelia bu	THN2_WHEAT
442	1	136	2.3	5	P32032	triticum ae	THNB_HORVU
443	1	136	2.3	5	P21742	hordeum vul	THNB_WHEAT
444	1	136	2.3	5	P01543	triticum ae	YIGA_SALTY
445	1	136	2.3	5	P55886	salmonella	COF1_DICDI
446	1	137	2.3	5	P54706	dictyostell	ERV4_YEAST
447	1	137	2.3	5	P53173	saccharomyc	KDGL_PSEDE
448	1	137	2.3	5	P29945	pseudomonas	RET4_HUMAN
449	1	137	2.3	5	P29373	homo sapien	RET4_MOUSE
450	1	137	2.3	5	P22935	mus musculu	THN3_HORVU
451	1	137	2.3	5	P08772	hordeum vul	THN6_HORVU
452	1	137	2.3	5	P09618	hordeum vul	ACP1_SPIOL
453	1	138	2.3	5	P07854	spinacia ol	CXA4_PIG
454	1	138	2.3	5	Q29559	sus scrofa	RET4_RAT
455	1	138	2.3	5	P51673	rattus norv	RL27_SOLTU
456	1	138	2.3	5	P41101	solanum tub	RM16_CHOCR
457	1	138	2.3	5	P48955	chondrus cr	RR11_ARATH
458	1	138	2.3	5	P56802	arabidopsis	RR11_PEA
459	1	138	2.3	5	P06587	pisum sativ	RR11_TOBAC
460	1	138	2.3	5	P06365	nicotiana t	UL73_HCMVA
461	1	138	2.3	5	P16795	human cytom	VL1_BPV6
462	1	138	2.3	5	P50807	bovine papi	Y087_HUMAN
463	1	138	2.3	5	Q14695	homo sapien	RK16_PORPU
464	1	139	2.3	5	P51307	porphyra pu	Y201_RICPR
465	1	139	2.3	5	Q9zdw4	rickettsia	Y28_BPT7
466	1	139	2.3	5	P03795	bacteriopha	YB04_FOWPM
467	1	140	2.3	5	P14362	fowlpox vir	YDS1_SCHPO
468	1	141	2.3	5	O14177	schizosacch	ATPE_HAEIN
469	1	142	2.3	5	P43718	haemophilus	RECA_STRCT
470	1	142	2.3	5	P41055	streptomyce	YWDC_BACSU
471	1	142	2.3	5	P39611	bacillus su	

Right column block:

No.		Len			Accession	Organism	ID
472	1	143	2.3	5	P27879	medicago sa	HS11_MEDSA
473	1	143	2.3	5	P01744	homo sapien	HV1C_HUMAN
474	1	143	2.3	5	Q9zlh3	helicobacte	RNH_HELPJ
475	1	143	2.3	5	P56120	helicobacte	RNH_HELPJ
476	1	143	2.3	5	P09561	zea mays (m	RR11_MAIZE
477	1	143	2.3	5	P12096	oryza sativ	RR11_ORYSA
478	1	143	2.3	5	Q10846	mycobacteri	YK11_MYCTU
479	1	143	2.3	5	Q43517	lycopersico	FER1_LYCES
480	1	144	2.3	5	P43313	helicobacte	NAPA_HELPY
481	1	144	2.3	5	P10769	cavia porce	NEU2_CAVPO
482	1	144	2.3	5	P44814	haemophilus	YIHZ_HAEIN
483	1	144	2.3	5	O76096	homo sapien	CYTF_HUMAN
484	1	145	2.3	5	062215	caenorhabdi	DHSD_CAEEL
485	1	145	2.3	5	083050	treponema p	Y004_TREPA
486	1	145	2.3	5	P32147	escherichia	YIHZ_ECOLI
487	1	146	2.3	5	P07429	xenopus tro	HBB1_XENTR
488	1	146	2.3	5	Q95189	gorilla gor	OB_GORGO
489	1	146	2.3	5	O02750	pan troglod	OB_PANTR
490	1	147	2.3	5	P07733	xenopus lae	GS17_XENLA
491	1	147	2.3	5	Q60440	mesocricetu	IL4_MESAU
492	1	147	2.3	5	P20096	rattus norv	IL4_RAT
493	1	147	2.3	5	P42067	medicago sa	MCP_MEDSA
494	1	147	2.3	5	P26947	bacillus fi	MUTX_BACFI
495	1	147	2.3	5	O67723	aquifex aeo	RS9_AQUAE
496	1	147	2.3	5	P19746	capripoxvir	VFP7_CAPVK
497	1	147	2.3	5	P24537	burkholderi	YI22_BURCE
498	1	148	2.3	5	P81583	cancer pagu	CPC4_CANPG
499	1	148	2.3	5	P55361	rhizobium s	Y4AN_RHISN
500	1	149	2.3	5	P24805	nicotiana t	TSJT_TOBAC
501	1	149	2.3	5	P27089	thiobacillu	YML3_THIFE
502	1	150	2.3	5	P27777	zea mays (m	FER1_MAIZE
503	1	150	2.3	5	Q48900	oryza sativ	HS11_ORYSA
504	1	151	2.3	5	Q49093	mycobacteri	CH60_MYCAV
505	1	151	2.3	5	Q49160	mycobacteri	CH60_MYCCH
506	1	151	2.3	5	Q50220	mycobacteri	CH60_MYCFO
507	1	151	2.3	5	P12810	mycobacteri	CH60_MYCMR
508	1	151	2.3	5	P54043	triticum ae	HS11_WHEAT
509	1	151	2.3	5	P46298	methanococc	RL19_METJA
510	1	151	2.3	5	O64246	pisum sativ	RS13_PEA
511	1	151	2.3	5	O14183	mycobacteri	VG55_BPMD2
512	1	151	2.3	5	P36108	schizosacch	YDS8_SCHPO
513	1	151	2.3	5	P43546	saccharomyc	YKA2_YEAST
514	1	152	2.3	5	P01161	rattus norv	ANF_RAT
515	1	152	2.3	5	P43546	saccharomyc	YFF7_YEAST
516	1	153	2.3	5	P30693	helianthus	HS11_HELAN
517	1	153	2.3	5	P02519	glycine max	HS11_SOYBN
518	1	153	2.3	5	P04793	glycine max	HS13_SOYBN
519	1	153	2.3	5	P01080	solanum tub	IP2K_SOLTU
520	1	153	2.3	5	P07632	rattus norv	SODC_RAT
521	1	154	2.3	5	P30734	micromys mi	VE6_MMPV
522	1	154	2.3	5	P45999	escherichia	FASC_ECOLI
523	1	154	2.3	5	P31673	oryza sativ	HS12_ORYSA
524	1	154	2.3	5	P13853	arabidopsis	HS12_ARATH
525	1	154	2.3	5	Q51506	glycine max	HS14_SOYBN
526	1	154	2.3	5	P27396	solanum tub	IP25_SOLTU
527	1	155	2.3	5	O32068	haloarcula	RL30_HALMA
528	1	155	2.3	5	Q95123	human papil	VE6_HPV66
529	1	156	2.3	5	P71512	chromatium	YPH7_CHRVI
530	1	156	2.3	5	P19036	arabidopsis	HS11_ARATH
531	1	156	2.3	5	P13853	arabidopsis	HS12_ARATH
532	1	157	2.3	5	Q51506	pseudomonas	SOXR_PSEAE
533	1	157	2.3	5	P27396	daucus caro	HS11_DAUCA
534	1	158	2.3	5	O32068	bacillus su	YT2F_BACSU
535	1	158	2.3	5	Q95123	bos taurus	DHSD_BOVIN
536	1	158	2.3	5	P71512	methylobact	HPPK_METEX
537	1	158	2.3	5	P19243	pisum sativ	HS11_PEA
538	1	158	2.3	5	P12994	pisum sativ	HS11_PEA
539	1	159	2.3	5	O05515	bacillus su	YDIB_BACSU
540	1	159	2.3	5	P11572	mycobacteri	19KD_MYCTU
541	1	159	2.3	5	P97175	mycobacteri	MP63_MYCTU
542	1	160	2.3	5	Q54313	treponema h	SMPA_TREHY
543	1	161	2.3	5	P39467	sulfolobus	RS4_SULAC
544	1	161	2.3	5	P46733	mycobacteri	19KD_MYCAV
544	1	161	2.3	5	Q05832	chenopodium	HS11_CHERU

Left block

Index			Accession	Organism	SeqID		Num		
545	5	2.3	P19037	arabidopsis	HS13_ARATH	1	161	2.3	5
546	5	2.3	P05478	glycine max	HS16_SOYBN	1	161	2.3	5
547	5	2.3	P53922	saccharomyc	YNM0_YEAST	1	161	2.3	5
548	5	2.3	P31502	mycobacteri	19KD_MYCIT	1	162	2.3	5
549	5	2.3	P48263	cyanophora	C550_CYAPA	1	162	2.3	5
550	5	2.3	O78454	guillardia	C550_GUITH	1	162	2.3	5
551	5	2.3	O60286	methanococc	Y227_METJA	1	162	2.3	5
552	5	2.3	P49510	odontella s	C550_ODOSI	1	163	2.3	5
553	5	2.3	P51199	porphyra pu	C550_PORPU	1	163	2.3	5
554	5	2.3	P45260	haemophilus	ILVH_HAEIN	1	163	2.3	5
555	5	2.3	P29921	paracoccus	NQO9_PARDE	1	163	2.3	5
556	5	2.3	P13333	bacteriopha	VG19_BPT4	1	163	2.3	5
557	5	2.3	Q38435	bacteriopha	VG19_BPT6	1	163	2.3	5
558	5	2.3	O57608	phasianus c	ING_PHACO	1	164	2.3	5
559	5	2.3	P25899	synechococc	PSAF_SYNEN	1	164	2.3	5
560	5	2.3	O69583	mycobacteri	QCRC_MYCLE	1	164	2.3	5
561	5	2.3	P34477	caenorhabdi	UBC7_CAEEL	1	164	2.3	5
562	5	2.3	P76510	escherichia	YFDN_ECOLI	1	164	2.3	5
563	5	2.3	P12061	salmonella	FM_SALEN	1	165	2.3	5
564	5	2.3	P55670	rhizobium s	Y4UD_RHISN	1	167	2.3	5
565	5	2.3	Q01545	pharbitis n	HS22_PHANI	1	167	2.3	5
566	5	2.3	P41159	homo sapien	OB_HUMAN	1	167	2.3	5
567	5	2.3	P42747	arabidopsis	UBC7_ARATH	1	167	2.3	5
568	5	2.3	P35434	rattus norv	ATPD_RAT	1	168	2.3	5
569	5	2.3	Q95136	bos taurus	DADR_BOVIN	1	168	2.3	5
570	5	2.3	P12448	human immun	REV_HV2SB	1	168	2.3	5
571	5	2.3	P12204	nicotiana t	YCF3_TOBAC	1	168	2.3	5
572	5	2.3	O78451	guillardia	ILVH_GUITH	1	169	2.3	5
573	5	2.3	Q46578	deinococcus	PLSX_DEIRA	1	169	2.3	5
574	5	2.3	Q92e22	rickettsia	RL10_RICPR	1	169	2.3	5
575	5	2.3	P52806	pinus thunb	YCF3_PINTH	1	169	2.3	5
576	5	2.3	Q00644	azospirillu	YNG2_AZOBR	1	169	2.3	5
577	5	2.3	P21316	escherichia	YR7B_ECOLI	1	169	2.3	5
578	5	2.3	P24515	drosophila	CH18_DROVI	1	170	2.3	5
579	5	2.3	Q99462	homo sapien	UBCG_HUMAN	1	170	2.3	5
580	5	2.3	P27324	zea mays (m	YCF3_MAIZE	1	170	2.3	5
581	5	2.3	P12203	oryza sativ	YCF3_ORYSA	1	170	2.3	5
582	5	2.3	P44975	haemophilus	LSPA_HAEIN	1	171	2.3	5
583	5	2.3	Q55141	synechocyst	ILVH_SYNY3	1	172	2.3	5
584	5	2.3	P81232	methanococc	Y21B_METJA	1	172	2.3	5
585	5	2.3	P42186	escherichia	PRSE_ECOLI	1	173	2.3	5
586	5	2.3	P25296	saccharomyc	CALB_YEAST	1	174	2.3	5
587	5	2.3	Q95137	bos taurus	DBDR_BOVIN	1	174	2.3	5
588	5	2.3	P55890	salmonella	DSBC_SALTY	1	174	2.3	5
589	5	2.3	P13421	serratia ma	FMA_SERMA	1	174	2.3	5
590	5	2.3	P20725	klebsiella	GSPB_KLEPN	1	174	2.3	5
591	5	2.3	P51230	porphyra pu	ILVH_PORPU	1	174	2.3	5
592	5	2.3	P18022	escherichia	SSBP_ECOLI	1	175	2.3	5
593	5	2.3	P28045	escherichia	SSBR_ECOLI	1	175	2.3	5
594	5	2.3	P04668	klebsiella	FLAW_KLEPN	1	175	2.3	5
595	5	2.3	P17288	escherichia	IPYR_ECOLI	1	175	2.3	5
596	5	2.3	P29941	pseudomonas	YCB8_PSEDE	1	176	2.3	5
597	5	2.3	P20462	macropus eu	LALP_MACEU	1	176	2.3	5
598	5	2.3	Q29144	trichosurus	LALP_TRIVU	1	176	2.3	5
599	5	2.3	P02390	escherichia	RL6_ECOLI	1	176	2.3	5
600	5	2.3	Q01223	vaccinia vi	V19R_VACCV	1	177	2.3	5
601	5	2.3	P46181	acyrthosiph	RL6_ACYKS	1	178	2.3	5
602	5	2.3	P15063	rattus norv	MYCB_RAT	1	178	2.3	5
603	5	2.3	P46898	bacillus su	RL6_BACSU	1	178	2.3	5
604	5	2.3	Q9zjs6	helicobacte	RL6_HELPJ	1	178	2.3	5
605	5	2.3	P56034	helicobacte	RL6_HELPY	1	178	2.3	5
606	5	2.3	P36251	citrus gree	RPOB_CGDAB	1	178	2.3	5
607	5	2.3	P18310	escherichia	SSBF_ECOLI	1	178	2.3	5
608	5	2.3	O84806	chlamydia t	PTH_CHLTR	1	179	2.3	5
609	5	2.3	O62771	monodelphis	SP17_MONDO	1	179	2.3	5
610	5	2.3	P18035	escherichia	TRBB_ECOLI	1	179	2.3	5
611	5	2.3	P47958	stochomys l	APT_STOLQ	1	180	2.3	5
612	5	2.3	P32889	homo sapien	ARF1_HUMAN	1	180	2.3	5
613	5	2.3	P51643	xenopus lae	ARF1_XENLA	1	180	2.3	5
614	5	2.3	P16587	homo sapien	ARF3_HUMAN	1	180	2.3	5
615	5	2.3	O02664	oryctolagus	DADR_RABIT	1	180	2.3	5
616	5	2.3	P01272	bos taurus	GLUC_BOVIN	1	180	2.3	5
617	5	2.3	P01273	mesocricetu	GLUC_MESAU	1	180	2.3	5

Right block

Index	Accession	Organism	SeqID		Num		
618	Q51893	prochloroco	Y52L_PROMA	1	180	2.3	5
619	P08655	staphylococ	YMER_STAAU	1	180	2.3	5
620	O69777	rhizobium e	YRP2_RHIET	1	180	2.3	5
621	P16500	mus musculu	ARF2_MOUSE	1	181	2.3	5
622	P35848	neurospora	OM20_NEUCR	1	181	2.3	5
623	P95987	sulfolobus	RS4_SULSO	1	181	2.3	5
624	P91924	dugesia jap	ARF_DUGJA	1	182	2.3	5
625	Q56224	thermus aqu	NQO9_THETH	1	182	2.3	5
626	P72179	paracoccus	SECY_PARDE	1	182	2.3	5
627	Q57457	haemophilus	Y922_HAEIN	1	182	2.3	5
628	P34383	caenorhabdi	YLP3_CAEEL	1	182	2.3	5
629	P21601	human cytom	IR14_HCMVA	1	183	2.3	5
630	P24577	burkholderi	YI71_BURCE	1	183	2.3	5
631	Q9z849	chlamydia p	GRPE_CHLPN	1	184	2.3	5
632	O51355	borrelia bu	NUSG_BORBU	1	184	2.3	5
633	Q24669	drosophila	TRSF_DROSI	1	184	2.3	5
634	P54139	mycobacteri	Y487_MYCLE	1	184	2.3	5
635	P15870	strongyloce	H1D_STRPU	1	185	2.3	5
636	P42764	atriplex ca	JIPH_ATRCA	1	185	2.3	5
637	P51193	porphyra pu	PSAF_PORPU	1	185	2.3	5
638	Q59448	francisella	SODC_FRATU	1	185	2.3	5
639	P40042	saccharomyc	YEP6_YEAST	1	185	2.3	5
640	P35750	sus scrofa	CAN1_PIG	1	186	2.3	5
641	P15454	saccharomyc	KGUA_YEAST	1	186	2.3	5
642	Q06877	photobacter	LUXP_PHOLE	1	186	2.3	5
643	P73376	synechocyst	Y50L_SYNY3	1	186	2.3	5
644	O56416	thermus aqu	YACE_THETH	1	186	2.3	5
645	O08545	mus musculu	EFA3_MOUSE	1	187	2.3	5
646	Q9x2h4	thermotoga	YI60_THEMA	1	187	2.3	5
647	Q9zdt7	rickettsia	EFP_RICPR	1	188	2.3	5
648	P15950	rattus norv	KLK3_RAT	1	188	2.3	5
649	P23593	rattus norv	APD_RAT	1	189	2.3	5
650	P10463	canis famil	CAYP_CANFA	1	189	2.3	5
651	P41150	oryctolagus	CAYP_RABIT	1	189	2.3	5
652	P30583	emericella	CHS1_EMENI	1	189	2.3	5
653	P49878	bos taurus	INAH_BOVIN	1	189	2.3	5
654	P28915	escherichia	YBFC_ECOLI	1	190	2.3	5
655	P77717	escherichia	YBAY_ECOLI	1	190	2.3	5
656	O22020	cyanidiosch	RR4_CYAME	1	191	2.3	5
657	Q09742	schizosacch	YB61_SCHPO	1	191	2.3	5
658	Q30236	glycine max	HS41_SOYBN	1	192	2.3	5
659	P15153	homo sapien	RAC2_HUMAN	1	192	2.3	5
660	Q05144	mus musculu	RAC2_MOUSE	1	192	2.3	5
661	O22029	cyanidium c	RR4_CYACA	1	192	2.3	5
662	P46587	candida alb	HS72_CANAL	1	193	2.3	5
663	P17543	escherichia	PAPJ_ECOLI	1	193	2.3	5
664	P42189	escherichia	PRSJ_ECOLI	1	193	2.3	5
665	Q9dw3	rickettsia	RNH2_RICPR	1	193	2.3	5
666	P44339	haemophilus	YB71_HAEIN	1	193	2.3	5
667	P54480	bacillus su	YQFW_BACSU	1	194	2.3	5
668	P43013	haemophilus	ACPD_HAEIN	1	194	2.3	5
669	P72736	synechocyst	HPPK_SYNY3	1	195	2.3	5
670	P28694	escherichia	MOG_ECOLI	1	195	2.3	5
671	O10101	schizosacch	RS7_SCHPO	1	195	2.3	5
672	Q39615	chlamydomon	PSAD_CHLRE	1	196	2.3	5
673	O67768	aquifex aeo	RNH2_AQUAE	1	196	2.3	5
674	Q46948	escherichia	THIJ_ECOLI	1	196	2.3	5
675	P19244	methanococc	TRAP_METJA	1	196	2.3	5
676	P44645	pisum sativ	HS41_PEA	1	197	2.3	5
677	P11596	haemophilus	MOG_HAEIN	1	197	2.3	5
678	P35809	drosophila	TRSF_DROME	1	197	2.3	5
679	O58292	schizophyll	XYNA_SCHCO	1	197	2.3	5
680	O58338	methanococc	Y882_METJA	1	197	2.3	5
681	O80188	methanococc	Y928_METJA	1	197	2.3	5
682	P17624	homo sapien	NGAL_HUMAN	1	198	2.3	5
683	P46492	emericella	NUDC_EMENI	1	198	2.3	5
684	O60232	haemophilus	YAEE_HAEIN	1	198	2.3	5
685	P20809	homo sapien	AA27_HUMAN	1	199	2.3	5
686	P47873	homo sapien	IL11_HUMAN	1	199	2.3	5
687	P26850	mus musculu	IL11_MOUSE	1	199	2.3	5
688	O84971	marchantia	NU6M_MARPO	1	199	2.3	5
689	Q17172	rhodobacter	NUOC_RHOCA	1	199	2.3	5
690	—	brugia mala	TDX2_BRUMA	1	199	2.3	5

Left column (entries 691–763). Constant columns across all rows: `1`, `2.3`, `5`.

No.	Grp	Name	Accession	Organism	1	2.3	5
691	200	NGAL_MOUSE	P11672	mus musculu	1	2.3	5
692	200	RS8_SCHPO	O14049	schizosacch	1	2.3	5
693	200	STRM_STRGR	P29783	streptomyce	1	2.3	5
694	200	YOR1_COYMV	P19200	commelina y	1	2.3	5
695	201	BLIP_STRCL	P35804	streptomyce	1	2.3	5
696	201	YBXB_BACSU	P37872	bacillus su	1	2.3	5
697	202	CUTI_BOTCI	Q00298	botrytis ci	1	2.3	5
698	202	ET1_BOVIN	P17322	bos taurus	1	2.3	5
699	202	ET1_RAT	P22388	rattus norv	1	2.3	5
700	202	HS2M_PEA	P46254	pisum sativ	1	2.3	5
701	202	RB18_LYMST	Q05976	lymnaea sta	1	2.3	5
702	202	YJ72_YEAST	P47142	saccharomyc	1	2.3	5
703	203	CAIE_ECOLI	P39206	escherichia	1	2.3	5
704	203	COBC_ECOLI	P52086	escherichia	1	2.3	5
705	203	GTS_OMMSL	P46088	ommastrephe	1	2.3	5
706	203	RAB7_DICDI	P36411	dictyosteli	1	2.3	5
707	204	TRPG_HALVO	P33974	halobacteri	1	2.3	5
708	205	DUT_RAT	P70583	rattus norv	1	2.3	5
709	205	RCP_VIBCH	P24550	vibrio chol	1	2.3	5
710	205	RS4_MYCGE	P47553	mycoplasma	1	2.3	5
711	205	RS4_MYCPN	P46775	mycoplasma	1	2.3	5
712	205	Y115_NPVOP	O10354	orgyia pseu	1	2.3	5
713	205	YETR_SHEPU	P46149	shewanella	1	2.3	5
714	206	R13A_PICMA	O65055	picea maria	1	2.3	5
715	206	RL3_THETH	P52860	thermus aqu	1	2.3	5
716	206	RNH2_METTH	O27102	methanobact	1	2.3	5
717	207	TRIC_RABIT	P02646	oryctolagus	1	2.3	5
718	207	HIS7_AZOBR	P18787	azospirillu	1	2.3	5
719	207	NU6M_PROWI	Q37626	prototheca	1	2.3	5
720	207	UPP_MYCTU	P94928	mycobacteri	1	2.3	5
721	208	HS26_DROME	P02517	drosophila	1	2.3	5
722	208	KGUA_MYCTU	P71659	mycobacteri	1	2.3	5
723	208	NU4M_MICPE	O21521	microtus pe	1	2.3	5
724	208	PSAD_LYCES	P12372	lycopersico	1	2.3	5
725	208	RL13_CHLSW	O48513	chlamydomon	1	2.3	5
726	209	PGSA_MYCTU	Q50611	mycobacteri	1	2.3	5
727	209	RNH2_HELPJ	Q9zrl1	helicobacte	1	2.3	5
728	209	RNH2_HELPY	P56121	helicobacte	1	2.3	5
729	210	YRBK_HAEIN	P45075	haemophilus	1	2.3	5
730	210	CUTI_PHYCP	P41754	phytophthor	1	2.3	5
731	211	ESAI_ERWST	P54656	erwinia ste	1	2.3	5
732	211	YG21_YEAST	P53251	saccharomyc	1	2.3	5
733	211	CCMA_PARDE	P52218	paracoccus	1	2.3	5
734	211	CRB3_CHICK	P55165	gallus gall	1	2.3	5
735	211	PMGY_SCHPO	P36623	schizosacch	1	2.3	5
736	211	RCSA_ERWAM	P20098	erwinia amy	1	2.3	5
737	211	RCSA_ERWST	P27488	erwinia ste	1	2.3	5
738	212	RS3A_ARCFU	O27964	archaeoglob	1	2.3	5
739	212	TER5_ECOLI	P21337	escherichia	1	2.3	5
740	212	PCPP_BPT4	P06807	bacteriopha	1	2.3	5
741	213	RL7_PICJA	P32101	pichia jadi	1	2.3	5
742	213	GPH_AQUAE	O67359	aquifex aeo	1	2.3	5
743	213	UBC_ASFB7	P25869	african swi	1	2.3	5
744	213	YJH2_YEAST	P48592	saccharomyc	1	2.3	5
745	214	RIR2_DROME	Q9z962	chlamydia p	1	2.3	5
746	214	RNH2_CHLPN	O58046	methanococc	1	2.3	5
747	214	Y629_METJA	O58117	methanococc	1	2.3	5
748	214	Y706_METJA	P34673	caenorhabdi	1	2.3	5
749	214	YO23_CAEEL	P26161	rhodobacter	1	2.3	5
750	215	YPU5_RHOCA	P27949	african swi	1	2.3	5
751	216	UBC_ASFB7	P50661	callimico g	1	2.3	5
752	216	OLIA_HUMAN	P47882	homo sapien	1	2.3	5
753	216	OLID_HUMAN	P47885	homo sapien	1	2.3	5
754	216	GTM1_HUMAN	P09488	homo sapien	1	2.3	5
755	217	GTM1_RAT	P04905	rattus norv	1	2.3	5
756	217	GTM2_HUMAN	P28161	rattus norv	1	2.3	5
757	217	GTM3_RAT	P08009	rattus norv	1	2.3	5
758	217	GTM5_HUMAN	P46439	homo sapien	1	2.3	5
759	217	GTMU_CAVPO	P16413	cavia porce	1	2.3	5
760	217	HNT2_YEAST	P49775	saccharomyc	1	2.3	5
761	217	RNH2_CHLTR	O84032	chlamydia t	1	2.3	5
762	217	RR3_PINTH	P41635	pinus thunb	1	2.3	5

Right column (entries 764–836). Constant columns across all rows: `1`, `2.3`, `5`.

No.	Grp	Name	Accession	Organism	1	2.3	5
764	217	YQIE_HAEIN	P44684	haemophilus	1	2.3	5
765	218	GTM4_HUMAN	Q03013	homo sapien	1	2.3	5
766	218	KAD_BORPE	P39068	bordetella	1	2.3	5
767	218	RR3_PICAB	O62951	picea ables	1	2.3	5
768	218	RS7_PYRHO	O59230	pyrococcus	1	2.3	5
769	218	YT44_CAEEL	Q11072	caenorhabdi	1	2.3	5
770	219	AROD_AQUAE	O66440	aquifex aeo	1	2.3	5
771	219	PIGF_HUMAN	Q07326	homo sapien	1	2.3	5
772	219	YJFK_ECOLI	P39293	escherichia	1	2.3	5
773	219	YLP4_ZYMMO	O66114	zymomonas m	1	2.3	5
774	220	GTC1_RAT	P04904	rattus norv	1	2.3	5
775	220	GTC2_RAT	P46418	rattus norv	1	2.3	5
776	220	GTC_MOUSE	P30115	mus musculu	1	2.3	5
777	220	NK14_MOUSE	P27814	mus musculu	1	2.3	5
778	220	PC83_SOLTU	P20347	solanum tub	1	2.3	5
779	220	RNH2_BRUME	O68821	brucella me	1	2.3	5
780	220	RNH2_PYRHO	O59351	pyrococcus	1	2.3	5
781	221	ATP6_CLOAB	O05097	clostridium	1	2.3	5
782	221	GTA1_PIG	P51781	sus scrofa	1	2.3	5
783	222	KADC_MAIZE	P43188	zea mays (m	1	2.3	5
784	222	UL20_HSV11	P10204	herpes simp	1	2.3	5
785	222	Y4YL_RHISN	P55720	rhizobium s	1	2.3	5
786	222	YKGH_ECOLI	P77180	escherichia	1	2.3	5
787	222	BIOD_MYCLE	P45486	mycobacteri	1	2.3	5
788	223	CYPH_NEUCR	P10255	neurospora	1	2.3	5
789	223	NARI_BACSU	P42177	bacillus su	1	2.3	5
790	223	PLF1_MOUSE	P04095	mus musculu	1	2.3	5
791	224	PLF2_MOUSE	P04768	mus musculu	1	2.3	5
792	224	PLF3_MOUSE	P18918	mus musculu	1	2.3	5
793	224	UL01_HCMVA	P16719	human cytom	1	2.3	5
794	224	UREG_MYCTU	P50051	mycobacteri	1	2.3	5
795	224	Y119_METJA	Q57583	methanococc	1	2.3	5
796	224	KDPE_ECOLI	P21866	escherichia	1	2.3	5
797	225	OL1G_HUMAN	P47887	homo sapien	1	2.3	5
798	225	BASP_BOVIN	P80724	bos taurus	1	2.3	5
799	226	C79A_HUMAN	P11912	homo sapien	1	2.3	5
800	226	LYCV_BPPHV	P13003	bacteriopha	1	2.3	5
801	226	YC65_MYCTU	Q11054	mycobacteri	1	2.3	5
802	226	DEFA_ANTMA	P23706	antirrhinum	1	2.3	5
803	227	EXOD_BPT4	P04536	bacteriopha	1	2.3	5
804	227	COX2_LUMTE	Q37545	lumbricus t	1	2.3	5
805	228	LPRH_MYCTU	P71687	mycobacteri	1	2.3	5
806	228	MENE_SALTY	P37418	salmonella	1	2.3	5
807	228	MP64_MYCTU	P19996	mycobacteri	1	2.3	5
808	228	RNH2_PYRKO	O74035	pyrococcus	1	2.3	5
809	228	YBFG_HAEIN	P44670	haemophilus	1	2.3	5
810	229	FLGF_RHIME	O54248	rhizobium m	1	2.3	5
811	229	NHB1_RHORH	P21220	rhodococcus	1	2.3	5
812	229	REPX_STAAU	P03862	staphylococ	1	2.3	5
813	229	VG01_BPML5	Q05218	mycobacteri	1	2.3	5
814	229	XYN1_TRIRE	P36218	trichoderma	1	2.3	5
815	230	CRP_RAT	P48199	rattus norv	1	2.3	5
816	230	D102_ARATH	Q05212	arabidopsis	1	2.3	5
817	230	RIBR_BACSU	P94465	bacillus su	1	2.3	5
818	230	RK1_ODOSI	P49544	odontella s	1	2.3	5
819	230	RNH2_METJA	Q57599	methanococc	1	2.3	5
820	230	SOML_HIPHI	P45641	hippoglossu	1	2.3	5
821	231	SOML_SOLSE	P45642	solea seneg	1	2.3	5
822	231	COX2_ALOPA	P98024	alouatta pa	1	2.3	5
823	231	COX2_LAGLA	P98036	lagothrix l	1	2.3	5
824	231	MAD1_PETHY	Q07472	petunia hyb	1	2.3	5
825	231	NFIL_PIG	P14057	sus scrofa	1	2.3	5
826	231	YZ82_MYCTU	P96864	mycobacteri	1	2.3	5
827	232	GU58_RAT	P35899	rattus norv	1	2.3	5
828	232	RSUA_HAEIN	P45124	haemophilus	1	2.3	5
829	232	YACM_BACSU	Q06755	bacillus su	1	2.3	5
830	232	YHCG_BACSU	P54591	bacillus su	1	2.3	5
831	233	YHAK_ECOLI	P42624	escherichia	1	2.3	5
832	233	YM27_MYCTU	Q10511	mycobacteri	1	2.3	5
833	233	COBC_SALTY	P39701	salmonella	1	2.3	5
834	234	DNAJ_RHILE	O33529	rhizobium l	1	2.3	5
835	234	Y014_SYNY3	Q57208	synechocyst	1	2.3	5
836	235	Y544_RICPR	Q9zd06	rickettsia	1	2.3	5

Index	Protein ID		Len			Accession	Organism
837	ARCA_HAEIN	1	236	2.3	5	P44918	haemophilus
838	CLCD_PSEPU	1	236	2.3	5	P11453	pseudomonas
839	GT6_SCHMA	1	236	2.3	5	P46435	schistosoma
840	RECO_HAEIN	1	236	2.3	5	P44642	haemophilus
841	ATPF_KLULA	1	237	2.3	5	O13349	kluyveromyc
842	RECA_NEIEG	1	237	2.3	5	P86382	neisseria e
843	RECA_NEILA	1	237	2.3	5	P97195	neisseria l
844	RECA_NEIPE	1	237	2.3	5	P86403	neisseria p
845	RECA_NEIPO	1	237	2.3	5	P96955	neisseria p
846	RECA_NEISI	1	237	2.3	5	P86411	neisseria s
847	RECA_NEISU	1	237	2.3	5	P86414	neisseria s
848	YS76_CAEEL	1	237	2.3	5	Q09619	caenorhabdi
849	EFA3_HUMAN	1	238	2.3	5	P52797	homo sapien
850	HS2C_WHEAT	1	238	2.3	5	Q00445	triticum ae
851	PRRC_RAT	1	238	2.3	5	P33579	rattus norv
852	RNH2_THEMA	1	238	2.3	5	Q9x017	thermotoga
853	TCBE_PSESQ	1	238	2.3	5	P27100	pseudomonas
854	UBC3_RABIT	1	238	2.3	5	Q29503	oryctolagus
855	Y554_AQUAE	1	238	2.3	5	O66829	aquifex aeo
856	APE1_AERPE	1	239	2.3	5	O73942	aeropyrum p
857	CLDE_HUMAN	1	239	2.3	5	O95500	homo sapien
858	CLDE_MOUSE	1	239	2.3	5	Q920s3	mus musculu
859	CYBH_RHILV	1	239	2.3	5	P27648	rhizobium l
860	GUNA_ASPAK	1	239	2.3	5	Q12679	aspergillus
861	KLK2_CAVPO	1	239	2.3	5	P12323	cavia porce
862	PNDC_ECOLI	1	239	2.3	5	P31215	escherichia
863	RL2_PYRHO	1	239	2.3	5	O59421	pyrococcus
864	UL20_HSVEB	1	239	2.3	5	P28971	equine herp
865	YQIK_BACSU	1	239	2.3	5	P54527	bacillus su
866	YWIC_BACSU	1	239	2.3	5	P46909	bacillus su
867	CD48_MOUSE	1	240	2.3	5	P18181	mus musculu
868	RNC_MYCTU	1	240	2.3	5	Q10962	mycobacteri
869	SUR1_FUGRU	1	240	2.3	5	O57593	fugu rubrip
870	YDGB_ECOLI	1	240	2.3	5	P52109	escherichia
871	BTR_BORPE	1	241	2.3	5	Q08530	bordetella
872	GLTL_ECOLI	1	241	2.3	5	P41076	escherichia
873	HUTC_KLEAE	1	241	2.3	5	P12380	klebsiella
874	TRPF_METTH	1	241	2.3	5	O27695	methanobact
875	YD96_YEAST	1	241	2.3	5	Q03941	saccharomyc
876	GLUA_CORGL	1	242	2.3	5	P48243	corynebacte
877	RECO_ECOLI	1	242	2.3	5	P15027	escherichia
878	RL2_METJA	1	242	2.3	5	P54017	methanococc
879	AGI6_TRYBB	1	243	2.3	5	Q26768	trypanosoma
880	TRY1_XENLA	1	243	2.3	5	P19799	xenopus lae
881	YN06_CAEEL	1	243	2.3	5	Q03615	caenorhabdi
882	FMCH_BACNO	1	244	2.3	5	P17419	bacteroides
883	KLKA_RAT	1	244	2.3	5	P36375	rattus norv
884	RDGA_ERWCA	1	244	2.3	5	Q47587	erwinia car
885	UPP_TOXGO	1	244	2.3	5	Q26998	toxoplasma
886	DCOP_SALTY	1	245	2.3	5	P07691	salmonella
887	LUXP_VIBHA	1	245	2.3	5	P54300	vibrio harv
888	PMX1_HUMAN	1	245	2.3	5	P54821	homo sapien
889	PMX1_MOUSE	1	245	2.3	5	P43271	mus musculu
890	WECG_ECOLI	1	246	2.3	5	P27836	escherichia
891	Y181_METJA	1	246	2.3	5	Q57640	methanococc
892	YP73_MYCTU	1	246	2.3	5	Q50648	mycobacteri
893	VPM_BPP2	1	247	2.3	5	P25476	bacteriopha
894	H1_PARAN	1	248	2.3	5	P02256	parechinus
895	PRC8_DICDI	1	248	2.3	5	Q27563	dictyostell
896	RSR1_CANAL	1	248	2.3	5	P52498	candida alb
897	CDSA_ECOLI	1	248	2.3	5	P06466	escherichia
898	HMX1_CHICK	1	249	2.3	5	P28361	gallus gall
899	RU2A_ARATH	1	249	2.3	5	P43333	arabidopsis
900	YE64_AQUAE	1	249	2.3	5	O67444	aquifex aeo
901	YS91_MYCTU	1	249	2.3	5	Q10812	mycobacteri
902	ODFP_HUMAN	1	250	2.3	5	Q14990	homo sapien
903	GLO2_ECOLI	1	251	2.3	5	Q47677	escherichia
904	TPIS_LEIME	1	251	2.3	5	P48499	leishmania
905	AG_BRANA	1	252	2.3	5	Q01540	brassica na
906	P29_MYCHR	1	252	2.3	5	P15361	mycoplasma
907	ADH1_DROHY	1	253	2.3	5	Q23236	drosophila
908	ADH1_DROMO	1	253	2.3	5	P09370	drosophila
909	ADH1_DROMT	1	253	2.3	5	P22246	drosophila
910	ADH1_DROMU	1	253	2.3	5	P07161	drosophila
911	ADH1_DRONA	1	253	2.3	5	P12854	drosophila
912	ADH2_DROBU	1	253	2.3	5	P25720	drosophila
913	ADH2_DROHY	1	253	2.3	5	P23237	drosophila
914	ADH2_DROMO	1	253	2.3	5	P07160	drosophila
915	ADH2_DROMY	1	253	2.3	5	P25721	drosophila
916	ADH2_DROWH	1	253	2.3	5	P24267	drosophila
917	ADH_DROAD	1	253	2.3	5	Q00669	drosophila
918	ADH_DRODI	1	253	2.3	5	P22245	drosophila
919	ADH_DROGU	1	253	2.3	5	Q09009	drosophila
920	ADH_DROHE	1	253	2.3	5	Q21898	drosophila
921	ADH_DROIM	1	253	2.3	5	Q07588	drosophila
922	ADH_DROMD	1	253	2.3	5	Q09010	drosophila
923	ADH_DROPE	1	253	2.3	5	P37473	drosophila
924	ADH_DROPL	1	253	2.3	5	P23277	drosophila
925	ADH_DROPS	1	253	2.3	5	P07158	drosophila
926	ADH_DROSL	1	253	2.3	5	P23278	drosophila
927	RM02_ACACA	1	253	2.3	5	P46763	acanthamoeb
928	SSRA_ARATH	1	253	2.3	5	P45434	arabidopsis
929	YMDA_CHLAU	1	253	2.3	5	O45826	chloroflexu
930	ADH_DROAM	1	254	2.3	5	P25139	drosophila
931	ADH_DROLE	1	254	2.3	5	P10807	drosophila
932	ADH_DROSU	1	254	2.3	5	Q03384	drosophila
933	MOTA_AQUAE	1	254	2.3	5	O67122	aquifex aeo
934	PRC6_YEAST	1	254	2.3	5	P40303	saccharomyc
935	YBFF_ECOLI	1	254	2.3	5	P75736	escherichia
936	YH11_MYCTU	1	254	2.3	5	O33210	mycobacteri
937	YN08_YEAST	1	254	2.3	5	P53904	saccharomyc
938	143P_ARATH	1	255	2.3	5	P42644	arabidopsis
939	ADH_DROER	1	255	2.3	5	P28483	drosophila
940	ADH_DROMA	1	255	2.3	5	P07162	drosophila
941	ADH_DROME	1	255	2.3	5	P00334	drosophila
942	ADH_DROOR	1	255	2.3	5	P07159	drosophila
943	ADH_DROSI	1	255	2.3	5	P07163	drosophila
944	ADH_DROTE	1	255	2.3	5	P28484	drosophila
945	ADH_DROYA	1	255	2.3	5	P26719	drosophila
946	CAPC_STAAU	1	255	2.3	5	P39852	staphylococ
947	STX6_HUMAN	1	255	2.3	5	O43752	homo sapien
948	STX6_RAT	1	255	2.3	5	Q63635	rattus norv
949	YHJH_ECOLI	1	255	2.3	5	Q37646	escherichia
950	ADH_DROTS	1	256	2.3	5	P51550	drosophila
951	ATPF_HUMAN	1	256	2.3	5	P24539	homo sapien
952	ATPF_RAT	1	256	2.3	5	P19511	rattus norv
953	PXBA_CANBO	1	256	2.3	5	Q00316	candida bol
954	PXBB_CANBO	1	256	2.3	5	Q00317	candida bol
955	Y107_NPVOP	1	256	2.3	5	O10346	orgyia pseu
956	YAFV_ECOLI	1	256	2.3	5	Q47679	escherichia
957	YH11_MYCLE	1	256	2.3	5	O05668	mycobacteri
958	ABCX_CYACA	1	257	2.3	5	P35020	cyanidium c
959	KLK1_MACFA	1	257	2.3	5	Q07276	macaca fasc
960	NOCP_AGRT5	1	257	2.3	5	P35116	agrobacteri
961	TDXM_MOUSE	1	257	2.3	5	P20108	mus musculu
962	Y036_MYCTU	1	257	2.3	5	P71606	mycobacteri
963	ARA5_ARATH	1	258	2.3	5	P28188	arabidopsis
964	CODY_BACSU	1	258	2.3	5	P39779	bacillus su
965	KLK1_PAPHA	1	258	2.3	5	Q28773	papio hamad
966	YJ98_MYCTU	1	258	2.3	5	Q10859	mycobacteri
967	YNW8_YEAST	1	258	2.3	5	P53862	saccharomyc
968	CAH2_BOVIN	1	259	2.3	5	P00921	bos taurus
969	CAH2_SHEEP	1	259	2.3	5	P00922	ovis aries
970	KLK2_RAT	1	259	2.3	5	P00759	rattus norv
971	KLK9_RAT	1	259	2.3	5	P07647	rattus norv
972	KLKB_RAT	1	259	2.3	5	P36376	rattus norv
973	KLKL_MOUSE	1	259	2.3	5	P15948	mus musculu
974	NODJ_RHILV	1	259	2.3	5	P06755	rhizobium l
975	PEB1_CAMJE	1	259	2.3	5	P45678	campylobact
976	RNPH_MYCLE	1	259	2.3	5	P37939	mycobacteri
977	SSAT_SALTY	1	259	2.3	5	P96068	salmonella
978	YBBM_ECOLI	1	259	2.3	5	P77307	escherichia
979	ARGT_SALTY	1	259	2.3	5	P02911	salmonella
980	CLCA_PSEPU	1	260	2.3	5	P11451	pseudomonas
981	COX3_PISOC	1	260	2.3	5	P25003	pisaster oc
982	NMA_HUMAN	1	260	2.3	5	Q13145	homo sapien

```
983     5   2.3   RFXK_HUMAN    260  1   O14593  homo sapien
984     5   2.3   CATG_HUMAN    261  1   P28293  mus musculu
985     5   2.3   COX3_HUMAN    261  1   P00414  homo sapien
986     5   2.3   KLK1_MOUSE    261  1   P15947  mus musculu
987     5   2.3   KLK3_MOUSE    261  1   P00756  mus musculu
988     5   2.3   KLK6_MOUSE    261  1   P00755  mus musculu
989     5   2.3   KLK7_RAT      261  1   P36373  rattus norv
990     5   2.3   KLK8_RAT      261  1   P36374  rattus norv
991     5   2.3   KLK9_MOUSE    261  1   P15949  mus musculu
992     5   2.3   KLKF_MOUSE    261  1   P04071  mus musculu
993     5   2.3   MOTB_BACSU    261  1   P28612  bacillus su
994     5   2.3   PRC9_HUMAN    261  1   P25789  homo sapien
995     5   2.3   PRC9_RAT      261  1   P21670  rattus norv
996     5   2.3   PROS_MACMU    261  1   P33619  macaca mula
997     5   2.3   UREH_HAEIN    261  1   P44397  haemophilus
998     5   2.3   KLK1_HUMAN    262  1   P06870  homo sapien
999     5   2.3   ODFP_BOVIN    262  1   Q29438  bos taurus
1000    5   2.3   ODFP_PIG      262  1   Q29077  sus scrofa
```

ALIGNMENTS

RESULT 1
YMFC_ECOLI
```
ID   YMFC_ECOLI    STANDARD;    PRT;   217 AA.
AC   P75966;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 24.9 KD PROTEIN IN TRMU-ICDA INTERGENIC REGION.
GN   YMFC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97061202.
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC       STRONG, TO H.INFLUENZAE HI0694.
```
```
DR   EMBL; AE000213; AAC74219.1; ALT_INIT.
DR   EMBL; D90748; BAA35957.1; -.
DR   EMBL; D90749; BAA35966.1; -.
DR   ECOGENE; EG13447; YMFC.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   217 AA;  24880 MW;  F7C7A7CEDC5FD3F6 CRC64;
```

```
Query Match             23.4%;  Score 52;  DB 1;  Length 217;
Best Local Similarity   100.0%;  Pred No. 1.2e-48;
Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    151 WLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMG 202
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 WLWPRNPPIRERKSIPTSWLKITLYEGRNRQVRRMTAHVGFPTLRLIRYAMG 200
```

RESULT 2
YMFC_HAEIN
```
ID   YMFC_HAEIN    STANDARD;    PRT;   240 AA.
AC   P44827;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0694.
GN   HI0694.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC       STRONG, TO E.COLI YMFC.
```
```
DR   EMBL; U32752; AAC22354.1; -.
DR   TIGR; HI0694; -.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   240 AA;  27464 MW;  0AD99EC61F52C01C CRC64;
```

```
Query Match              6.8%;  Score 15;  DB 1;  Length 240;
Best Local Similarity   100.0%;  Pred. No. 1.6e-08;
Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     74 VYAAGRLDRDSEGLL 88
          ||||||||||||||
Db     84 VYAAGRLDRDSEGLL 98
```

RESULT 3
Y612_SYNY3
```
ID   Y612_SYNY3    STANDARD;    PRT;   261 AA.
AC   P72581; -
```

```
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL 21.0 KD PROTEIN SLR0612.
GN   SLR0612.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97061201.
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
RN   [2]
RP   CONCEPTUAL TRANSLATION.
RA   Bairoch A.;
RL   Unpublished observations (SEP-1998).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: A FRAMESHIFT WAS
CC       INTRODUCED IN POSITION 20 TO PRODUCE THIS ORF.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D90899; BAA16580.1; ALT_FRAME.
DR   PROSITE; PS01149; PSI_RSU; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 261 AA; 29561 MW; 601A453085C04A69 CRC64;

Query Match             5.4%;  Score 12;  DB 1;  Length 261;
  Best Local Similarity 100.0%;  Pred. No. 3e-05;
  Matches   12;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy  176 EGRNRQVRMTA 187
         |||||||||||
Db  175 EGRNRQVRMTA 186

RESULT 4
RLUB_BACSU
ID   RLUB_BACSU   STANDARD;      PRT;   229 AA.
AC   P35159;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RIBOSOMAL LARGE SUBUNIT PSEUDOURIDINE SYNTHASE B (EC 4.2.1.70)
DE   (PSEUDOURIDYLATE SYNTHASE) (URACIL HYDROLYASE).
GN   RLUB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE; 95020538.
RA   Sorokin A.V., Zumstein E., Azevedo V., Ehrlich S.D., Serror P.;
RT   "The organization of the Bacillus subtilis 168 chromosome region
RT   between the spoVA and serA genetic loci, based on sequence data.";
RL   Mol. Microbiol. 10:385-395(1993).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE; 98325071.
RA   Niu L., Ofengand J.;
RL   Unpublished results, cited by:
RL   Conrad J., Sun D., Englund N., Ofengand J.;
RL   J. Biol. Chem. 273:18562-18566(1998).
CC   -!- FUNCTION: RESPONSIBLE FOR SYNTHESIS OF PSEUDOURIDINE FROM URACIL-
CC       2650 IN 23S RIBOSOMAL RNA.
CC   -!- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE
CC       5'-PHOSPHATE + H(2)O.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF PSEUDOURIDINE SYNTHASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L09228; AAA67493.1; -.
DR   EMBL; Z99116; CAB14248.1; -.
DR   PIR; S45555; S45555.
DR   SUBTILIST; BG10530; RLUB.
DR   PROSITE; PS01149; PSI_RSU; 1.
DR   PFAM; PF00849; YABO; 1.
DR   PFAM; PF01479; S4; 1.
KW   Lyase.
SQ   SEQUENCE 229 AA; 26025 MW; 0CD54C2EE7A5ADE8 CRC64;

Query Match             4.5%;  Score 10;  DB 1;  Length 229;
  Best Local Similarity 100.0%;  Pred. No. 0.0039;
  Matches   10;  Conservative   10;  Mismatches   0;  Indels    0;  Gaps    0;

Qy  176 EGRNRQVRRM 185
         ||||||||||
Db  184 EGRNRQVRRM 193

RESULT 5
CH60_RICTS
ID   CH60_RICTS   STANDARD;      PRT;   555 AA.
AC   P16625;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-FEB-2000 (Rel. 39, Last annotation update)
DE   60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (MAJOR ANTIGEN 58)
DE   (58 KD ANTIGEN).
GN   MOPA OR GROEL OR STA58.
OS   Rickettsia tsutsugamushi.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Orientia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90216005.
RA   Stover C.K., Marana D.P., Dasch G.A., Oaks E.V.;
RT   "Molecular cloning and sequence analysis of the sta58 major antigen
RT   gene of Rickettsia tsutsugamushi: sequence homology and antigenic
RT   comparison of Sta58 to the 60-kilodalton family of stress proteins.";
RL   Infect. Immun. 58:1360-1368(1990).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M31887; AAA26393.1; -.
DR   PIR; B41492; B41492.
DR   HSSP; P06139; 1GRL.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR   PFAM; PF00118; cpn60_TCP1; 1.
KW   Chaperone; ATP-binding; Antigen.
SQ   SEQUENCE   555 AA;  59729 MW;  16400249D3FC03B4 CRC64;

Query Match             3.6%;  Score 8;  DB 1;  Length 555;
Best Local Similarity   100.0%;  Pred. No. 1.3;
Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      124 LAKLRNGV 131
            ||||||||
Db      372 LAKLRNGV 379


RESULT 6
REPA_AGRTU
ID   REPA_AGRTU     STANDARD;      PRT;    250 AA.
AC   P15394;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   REPLICATING PROTEIN.
GN   REPA.
OS   Agrobacterium tumefaciens.
OG   Plasmid pTAR.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Agrobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88257036.
RA   Gallie D.R., Kado C.I.;
RT   "Minimal region necessary for autonomous replication of pTAR.";
RL   J. Bacteriol. 170:3170-3176(1988).
CC   -!- FUNCTION: REQUIRED FOR REPLICATION. IT LIKELY REGULATES PTAR
CC       COPY NUMBER.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M21299; AAD15307.1; -.
DR   PIR; A43662; A43662.
KW   Plasmid; DNA replication.
SQ   SEQUENCE   250 AA;  27987 MW;  CCED10653483I979 CRC64;

Query Match             3.2%;  Score 7;  DB 1;  Length 250;
Best Local Similarity   100.0%;  Pred. No. 7.4;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      102 QPGKRTG 108
            |||||||
Db       59 QPGKRTG 65


RESULT 7
STCV_EMENI
ID   STCV_EMENI     STANDARD;      PRT;    387 AA.
AC   Q00727;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV
DE   (EC 1.1.1.-).
GN   STCV.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Plectomycetes; Eurotiales;
OC   Trichocomaceae; Emericella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
RX   MEDLINE; 96202293.
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF 5'-
CC       HYDROXYAVERANTIN TO FORM AVERUFIN.
CC   -!- PATHWAY: STERIGMATOCYSTIN BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U34740; AAC49206.1; -.
KW   Oxidoreductase.
FT   ACT_SITE    148    148     HYDROGEN-BOND DONOR (POTENTIAL).
SQ   SEQUENCE   387 AA;  43548 MW;  1621588273B85E8C CRC64;

Query Match             3.2%;  Score 7;  DB 1;  Length 387;
Best Local Similarity   100.0%;  Pred. No. 11;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      122 ASLAKLR 128
            |||||||
Db      132 ASLAKLR 138


RESULT 8
NFM_PIG
ID   NFM_PIG        STANDARD;      PRT;    454 AA.
AC   P08552;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M)
DE   (FRAGMENT).
GN   NEFM.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SPINAL CORD;
RX   MEDLINE; 85076594.
RA   Geisler N., Fischer S., Vandekerckhove J., Plessmann U., Weber K.;
RT   "Hybrid character of a large neurofilament protein (NF-M):
RT   intermediate filament type sequence followed by a long and acidic
RT   carboxy-terminal extension.";
RL   EMBO J. 3:2701-2706(1984).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
```

```
CC  -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC      OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC      LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC      COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC  -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR  PIR; A05075; QFPGM.
DR  PROSITE; PS00226; IF; 1.
DR  PFAM; PF00038; filament; 1.
KW  Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW  Acetylation; Phosphorylation; Glycoprotein.
FT  MOD_RES        1              1       ACETYLATION.
FT  DOMAIN         1            104       HEAD.
FT  DOMAIN       105            412       ROD.
FT  DOMAIN       413           >454       TAIL.
FT  DOMAIN       105            136       COIL 1A.
FT  DOMAIN       137            149       LINKER 1.
FT  DOMAIN       150            248       COIL 1B.
FT  DOMAIN       249            265       LINKER 12.
FT  DOMAIN       266            287       COIL 2A.
FT  DOMAIN       288            291       LINKER 2.
FT  DOMAIN       292            411       COIL 2B.
FT  CARBOHYD      47             47       O-LINKED (GLCNAC) (BY SIMILARITY).
FT  CARBOHYD     432            432       O-LINKED (GLCNAC) (BY SIMILARITY).
FT  NON_TER      454            454
SQ  SEQUENCE     454 AA;    51854 MW;   C809282EEC903A6B CRC64;

Query Match             3.2%;    Score 7;    DB 1;   Length 454;
Best Local Similarity 100.0%;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
Matches 7;

Qy    203  SYTLDSL 209
           |||||||
Db      1  SYTLDSL 7

RESULT 9
PDI_ASPNG
ID  PDI_ASPNG      STANDARD;      PRT;    515 AA.
AC  Q12730;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-FEB-2000 (Rel. 39, Last annotation update)
DE  PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (EC 5.3.4.1).
GN  PDIA OR PDI1.
OS  Aspergillus niger.
OC  Eukaryota; Fungi; Ascomycota; Plectomycetes; Eurotiales;
OC  Trichocomaceae; anamorphic Trichocomaceae; Aspergillus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CBS 120.49 / N400;
RX  MEDLINE; 97174110.
RA  Ngiam C., Jeenes D.J., Archer D.B.;
RT  "Isolation and characterisation of a gene encoding protein disulphide
RT  isomerase, pdiA, from Aspergillus niger.";
RL  Curr. Genet. 31:133-138(1997).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=NRRL 3;
RA  Malpricht S.;
RL  Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC  -!- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
CC      DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
CC      HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
CC  -!- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
CC      DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
CC  -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).
CC  -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is .in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; X98797; CAA67332.1; -.
DR  EMBL; X89449; CAA61619.1; -.
DR  HSSP; P07237; 1MEK.
DR  PRINTS; PR00421; THIOREDOXIN.
DR  PROSITE; PS00014; ER_TARGET; 1.
DR  PROSITE; PS00194; THIOREDOXIN; 2.
DR  PFAM; PF00085; thiored; 3.
KW  Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT  SIGNAL         1             20       POTENTIAL.
FT  CHAIN         21            515       PROTEIN DISULFIDE ISOMERASE.
FT  DISULFID      54             57       REDOX-ACTIVE (BY SIMILARITY).
FT  DISULFID     389            392       REDOX-ACTIVE (BY SIMILARITY).
FT  SITE         512            515       PREVENT SECRETION FROM ER (POTENTIAL).
SQ  SEQUENCE     515 AA;    56291 MW;   2B0058B788400AD9 CRC64;

Query Match             3.2%;    Score 7;    DB 1;   Length 515;
Best Local Similarity 100.0%;  Pred. No. 14;
Matches 7;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    190  GFPTLRL 196
           |||||||
Db    434  GFPTLRL 440

RESULT 10
TCPG_YEAST
ID  TCPG_YEAST     STANDARD;      PRT;    534 AA.
AC  P39077;
DT  01-FEB-1995 (Rel. 31, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  T-COMPLEX PROTEIN 1, GAMMA SUBUNIT (TCP-1-GAMMA) (CCT-GAMMA).
GN  CCT3 OR TCP3 OR BIN2 OR YJL014W OR J1336.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC  Saccharomycetaceae; Saccharomyces.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 94377502.
RA  Chen X., Sullivan D.S., Huffaker T.C.;
RT  "Two yeast genes with similarity to TCP-1 are required for
RT  microtubule and actin function in vivo.";
RL  Proc. Natl. Acad. Sci. U.S.A. 91:9111-9115(1994).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  To Van D., Perea J., Jacq C.;
RL  Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC  -!- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
CC      ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC      ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
CC      FORMATION.
CC  -!- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KD THAT
CC      FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
CC  -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC  -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; U09480; AAA21658.1; -.
DR  EMBL; Z49289; CAA89305.1; -.
DR  PIR; S48875; S48875.
DR  HSSP; P48424; 1ASX.
```

```
DR   SGD; L0002270; CCT3.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00750; TCP1_1; 1.
DR   PROSITE; PS00751; TCP1_2; 1.
DR   PROSITE; PS00995; TCP1_3; 1.
DR   PFAM; PF00118; cpn60_TCP1; 1.
KW   Chaperone; ATP-binding; Multigene family.
FT   CONFLICT     25     26      IT -> HA (IN REF. 1).
FT   CONFLICT     86     89      DEEV -> ERRG (IN REF. 1).
FT   CONFLICT    111    111      L -> F (IN REF. 1).
FT   CONFLICT    265    265      K -> M (IN REF. 1).
FT   CONFLICT    274    274      Q -> L (IN REF. 1).
FT   CONFLICT    292    292      MISSING (IN REF. 1).
FT   CONFLICT    478    478      G -> A (IN REF. 1).
SQ   SEQUENCE   534 AA;  58814 MW;  5C5AFF3D67D9A3B6 CRC64;

Query Match             3.2%;  Score 7;  DB 1;  Length 534;
Best Local Similarity   100.0%;  Pred. No. 15;
Matches     7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    88 LVLTNDG 94
         |||||||
Db    55 LVLTNDG 61

RESULT  11
TCP2_YEAST
ID   TCP2_YEAST     STANDARD;     PRT;   546 AA.
AC   P39079;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   T-COMPLEX PROTEIN 1, ZETA SUBUNIT (TCP-1-ZETA) (CCT-ZETA).
GN   CCT6 OR TCP6 OR TCP20 OR YDR188W OR YDR395.21.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94308102.
RA   Li W.-Z., Lin P., Frydman J., Boal T.R., Cardillo T.S., Richard L.M.,
RA   Toth D., Lichtman M.A., Hartl F.-U., Sherman F., Segel G.B.,
RT   "Tcp20, a subunit of the eukaryotic TRiC chaperonin from humans and
RT   yeast.";
RL   J. Biol. Chem. 269:18616-18622(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Murphy L., Harris D.E., Barrell B.G., Rajandream M.A.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
CC       ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC       ACTIN AND TUBULIN. IN YEAST MAY PLAY A ROLE IN MITOTIC SPINDLE
CC       FORMATION.
CC   -!- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KD THAT
CC       FORMS TWO STACKED RINGS; 12 TO 16 NM IN DIAMETER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
```

```
DR   EMBL; L27698; AAA35140.1; -.
DR   EMBL; Z46727; CAA86694.1; -.
DR   PIR; S48086; S48086.
DR   HSSP; P48424; 1ASX.
DR   YEPD; 6690; -.
```

```
DR   SGD; L0002271; CCT6.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00750; TCP1_1; 1.
DR   PROSITE; PS00751; TCP1_2; 1.
DR   PROSITE; PS00995; TCP1_3; 1.
DR   PFAM; PF00118; cpn60_TCP1; 1.
KW   Chaperone; ATP-binding; Multigene family.
SQ   SEQUENCE   546 AA;  59923 MW;  7D2A2AB4526D3DB4 CRC64;

Query Match             3.2%;  Score 7;  DB 1;  Length 546;
Best Local Similarity   100.0%;  Pred. No. 15;
Matches     7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    60 AGRSTLK 66
         |||||||
Db   536 AGRSTLK 542

RESULT  12
SP20_YEAST
ID   SP20_YEAST     STANDARD;     PRT;   604 AA.
AC   P50875;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TRANSCRIPTION FACTOR SPT20.
GN   SPT20 OR ADA5 OR YOL148C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Roberts S.M., Winston F.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BWG1-7A;
RA   Marcus G.A., Silverman N., Guarente L.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Arino J., Casamayor A., Gamo F.J., Gancedo C., Lafuente M.J.,
RA   Aldea M., Casas C., Herrero E.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 50-604 FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE; 96132030.
RA   Casamayor A., Aldea M., Casas C., Herrero E., Gamo F.J.,
RA   Lafuente M.J., Gancedo C., Arino J.;
RT   "DNA sequence analysis of a 13 kbp fragment of the left arm of yeast
RT   chromosome XV containing seven new open reading frames.";
RL   Yeast 11:1281-1288(1995).
CC   -!- FUNCTION: HFI1/ADA1 AND SPT20/ADA5 MAY RECRUIT TATA BINDING
CC       PROTEIN (TBP) AND POSSIBLY OTHER BASAL FACTORS TO BIND TO THE TATA
CC       BOX WHILE ADA2, ADA3 AND GCN5 FUNCTION TO ACETYLATE NUCLEOSOMES
CC       OPENING UP THE PROMOTER REGION.
CC   -!- SUBUNIT: PART OF THE ADA/GCN5 COMPLEX THAT CONSISTS OF HFI1/ADA1,
CC       ADA2, ADA3, SPT20/ADA5 AND GCN5.
```

```
DR   EMBL; U22063; AAB07900.1; -.
DR   EMBL; U43153; AAB07899.1; -.
DR   EMBL; Z74890; CAA99169.1; -.
```

DR   EMBL; Z48239; CAB88279.1; -.
DR   SGD; L0002593; SPT20.
KW   Transcription regulation.
FT   DOMAIN   157   162   POLY-GLN.
FT   DOMAIN   235   240   POLY-SER.
FT   DOMAIN   422   425   POLY-SER.
FT   DOMAIN   454   463   POLY-ALA.
FT   DOMAIN   552   559   POLY-ASN.
FT   CONFLICT   293   293   Y -> S (IN REF. 4).
SQ   SEQUENCE   604 AA;   67796 MW; 3D67937B65F9AA6A CRC64;

Query Match   3.2%; Score 7; DB 1; Length 604;
Best Local Similarity 100.0%; Pred. No.17;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   216 DVTPKEN 222
            |||||||
Db   223 DVTPKEN 229

RESULT 13
COG9_RABIT
ID   COG9_RABIT   STANDARD;   PRT;   707 AA.
AC   P41246;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   92 KD TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.35) (92 KD GELATINASE)
DE   (MATRIX METALLOPROTEINASE-9) (MMP-9) (GELATINASE B).
DE   MMP9.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAPANESE WHITE; TISSUE=BONE;
RX   MEDLINE; 94253056.
RA   Tezuka K.I., Nemoto K., Tezuka Y., Sato T., Ikeda Y., Kobori M.,
RA   Kawashima H., Eguchi H., Hakeda Y., Kumegawa M.;
RT   "Identification of matrix metalloproteinase 9 in rabbit osteoclasts.";
RL   J. Biol. Chem. 269:15006-15009(1994).
RN   [2]
RP   SEQUENCE OF 1-171 FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=LIVER;
RX   MEDLINE; 95050662.
RA   Fini M.E., Bartlett J.D., Matsubara M., Rinehart W.B., Mody M.K.;
RA   Girard M.T., Rainville M.;
RT   "The rabbit gene for 92-kDa matrix metalloproteinase. Role of AP1 and
RT   AP2 in cell type-specific transcription.";
RL   J. Biol. Chem. 269:28620-28628(1994).
CC   -!- FUNCTION: COULD PLAY A ROLE IN BONE OSTEOCLASTIC RESORPTION.
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF GELATIN TYPES I AND V AND COLLAGEN
CC   TYPES IV AND V.
CC   -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC   -!- TISSUE SPECIFICITY: OSTEOCLASTS.
CC   -!- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II
CC   MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN
CC   BINDS GELATIN.
CC   -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE II-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC   METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D26514; BAA05520.1; -.

DR   EMBL; L36050; AAA64358.1; -.
DR   HSSP; P08254; 1SLN.
DR   PRINTS; PR00013; FNTYPEII.
DR   PRINTS; PR00138; MATRIXIN.
DR   PROSITE; PS00023; FIBRONECTIN_2; 3.
DR   PROSITE; PS00024; HEMOPEXIN; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR   PFAM; PF00040; fn2; 3.
DR   PFAM; PF00045; hemopexin; 4.
DR   PFAM; PF00413; Peptidase_M10.1.
KW   Hydrolase; Metalloprotease; Glycoprotein; zinc; Zymogen; Calcium;
KW   Collagen degradation; Extracellular matrix; Signal.
FT   SIGNAL   1   19   BY SIMILARITY.
FT   PROPEP   20   106   ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN   107   707   92 KD TYPE IV COLLAGENASE.
FT   DOMAIN   107   104   AUTOINHIBITOR REGION.
FT   DOMAIN   223   280   FIBRONECTIN TYPE-II.
FT   DOMAIN   281   339   FIBRONECTIN TYPE-II.
FT   DOMAIN   340   397   FIBRONECTIN TYPE-II.
FT   DOMAIN   511   707   HEMOPEXIN-LIKE.
FT   METAL   401   401   ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   402   402   BY SIMILARITY.
FT   METAL   405   405   ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL   411   411   ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD   88   88   POTENTIAL.
FT   CARBOHYD   120   120   POTENTIAL.
FT   CARBOHYD   127   127   POTENTIAL.
FT   DISULFID   516   704   BY SIMILARITY.
FT   CONFLICT   76   76   K -> P (IN REF. 2).
FT   CONFLICT   100   102   GVP -> ASR (IN REF. 2).
SQ   SEQUENCE   707 AA;   78307 MW; 053BCE8DC4D4758F CRC64;

Query Match   3.2%; Score 7; DB 1; Length 707;
Best Local Similarity 100.0%; Pred. No. 19;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   35 TPEPQPT 41
            |||||||
Db   456 TPEPQPT 462

RESULT 14
ACOC_CUCMC
ID   ACOC_CUCMC   STANDARD;   PRT;   764 AA.
AC   Q42669;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ACONITASE (EC 4.2.1.3) (ACONITATE HYDRATASE) (CITRATE HYDRO-LYASE)
DE   (FRAGMENT).
GN   ACO.
OS   Cucumis melo var. conomon (Oriental pickling melon).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;
OC   Cucumis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CANTALOUPE; TISSUE=FRUIT;
RX   MEDLINE; 95229629.
RA   Peyret P., Perez P., Alric M.;
RT   "Structure, genomic organization, and expression of the Arabidopsis
RT   thaliana aconitase gene. Plant aconitase show significant homology
RT   with mammalian iron-responsive element-binding protein.";
RL   J. Biol. Chem. 270:8131-8137(1995).
CC   -!- CATALYTIC ACTIVITY: CITRATE = CIS-ACONITATE + H(2)O.
CC   -!- PATHWAY: GLYOXYLATE BYPASS, PLAYS A ROLE IN GLUCONEOGENESIS FROM
CC   STORED OIL.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE
CC   (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE

```
CC    CATALYTIC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND
CC    ISOCITRATE.
CC -!- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC --------------------------------------------------------------------------
DR   EMBL; X82840; CAA58047.1; -.
DR   HSSP; P20004; 1FGH.
DR   PROSITE; PS00450; ACONITASE_1; 1.
DR   PROSITE; PS01244; ACONITASE_2; 1.
DR   PFAM; PF00330; aconitase; 1.
DR   PFAM; PF00694; Aconitase_C; 1.
KW   Lyase; Glyoxylate bypass; Iron-sulfur; 4Fe-4S.
FT   NON_TER       1      1
FT   METAL       307    307       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       372    372       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       375    375       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   764 AA;  83273 MW;  E4A9B011FC6922F0 CRC64;

Query Match             3.28;  Score 7;  DB 1;  Length 764;
Best Local Similarity 100.0%;         Pred. No. 21;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        126 KLRNGVT 132
              |||||||
Db        120 KLRNGVT 126

RESULT 15
NFM_RAT
ID   NFM_RAT        STANDARD;      PRT;   845 AA.
AC   P12839;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M).
GN   NFM OR NFM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87282618.
RA   Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT   "Complete amino acid sequence and in vitro expression of rat NF-M,
RT   the middle molecular weight neurofilament protein.";
RL   J. Neurosci. 7:2590-2599(1987).
RN   [2]
RP   PHOSPHORYLATION SITES, AND REVISION TO 500.
RX   MEDLINE; 92165797.
RA   Xu Z.-S., Liu W.-S., Willard M.B.;
RT   "Identification of six phosphorylation sites in the COOH-terminal
RT   tail region of the rat neurofilament protein M.";
RL   J. Biol. Chem. 267:4467-4471(1992).
RN   [3]
RP   CARBOHYDRATE-BINDING SITES.
RX   MEDLINE; 9334642I.
RA   Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA   Hart G.W.;
RT   "Glycosylation of mammalian neurofilaments. Localization of multiple
RT   o-linked N-acetylglucosamine moieties on neurofilament polypeptides
RT   L and M.";
RL   J. Biol. Chem. 268:16679-16687(1993).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC    AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC    PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC    THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC    INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC    OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC    OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC    LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC    COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC --------------------------------------------------------------------------
DR   EMBL; M18628; AAA41696.1; -.
DR   PROSITE; PS00226; IF; 1.
DR   PFAM; PF00038; filament; 1.
KW   Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW   Phosphorylation; Glycoprotein.
FT   INIT_MET      0      0
FT   DOMAIN        1    103       BY SIMILARITY.
FT   DOMAIN      103    410       HEAD.
FT   DOMAIN      411    844       TAIL.
FT   DOMAIN      103    134       ROD.
FT   DOMAIN      135    147       COIL 1A.
FT   DOMAIN      148    246       LINKER 1.
FT   DOMAIN      247    263       COIL 1B.
FT   DOMAIN      264    285       LINKER 12.
FT   DOMAIN      286    289       COIL 2A.
FT   DOMAIN      290    410       LINKER 2.
FT   DOMAIN              410      COIL 2B.
FT   CARBOHYD     47     47       O-LINKED (GLCNAC).
FT   CARBOHYD    430    430       O-LINKED (GLCNAC).
FT   MOD_RES     502    502       PHOSPHORYLATION.
FT   MOD_RES     506    506       PHOSPHORYLATION.
FT   MOD_RES     536    536       PHOSPHORYLATION.
FT   MOD_RES     603    603       PHOSPHORYLATION.
FT   MOD_RES     608    608       PHOSPHORYLATION.
FT   MOD_RES     666    666       PHOSPHORYLATION.
FT   CONFLICT    500    500       MISSING (IN REF. 1).
SQ   SEQUENCE   845 AA;  95660 MW;  316C41655B11197D CRC64;

Query Match             3.28;  Score 7;  DB 1;  Length 845;
Best Local Similarity 100.0%;         Pred. No. 23;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        203 SYTLDSL 209
              |||||||
Db          1 SYTLDSL 7

Search completed: June 10, 2000, 11:49:39
Job time: 1350 sec
```

us-09-252-691-7056.oligo.rsp

```
                    GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June 10, 2000, 11:25:05 ; Search time 16.42 Seconds
                                        (without alignments)
                                        937.404 Million cell updates/sec

Title:            US-09-252-691-7056
Perfect score:    222
Sequence:         1 AIMRQLITPENTMKTISFRK.........SYTLDSLANGEWRDVTPKEN 222

Scoring table:    OLIGO
                  Gapop 60.0 , Gapext 60.0
Searched:         225878 seqs, 69334122 residues
Word size :       0

Total number of hits satisfying chosen parameters:   225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 1000 summaries

Database :  SPTREMBL_12:*
         1:  sp_archaea:*
         2:  sp_bacteria:*
         3:  sp_fungi:*
         4:  sp_human:*
         5:  sp_invertebrate:*
         6:  sp_mammal:*
         7:  sp_mhc:*
         8:  sp_organelle:*
         9:  sp_phage:*
        10:  sp_plant:*
        11:  sp_rodent:*
        12:  sp_virus:*
        13:  sp_vertebrate:*
        14:  sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	3.6	341	5	P90743	caenorhabdi
2	8	3.6	448	2	O83284	treponema p
3	8	3.6	1494	5	Q20943	caenorhabdi
4	8	3.6	1525	9	Q94137	caenorhabdi
5	7	3.2	66	2	Q38044	bacteriopha
6	7	3.2	288	2	Q51543	pseudomonas
7	7	3.2	312	2	P72806	synechocyst
8	7	3.2	352	2	O05938	pseudomonas
9	7	3.2	356	5	Q9Y179	trichomonas
10	7	3.2	406	2	Q92MQ3	helicobacte
11	7	3.2	416	3	O00206	aspergillus
12	7	3.2	444	2	O86823	streptomyce
13	7	3.2	514	2	O25572	helicobacte
14	7	3.2	514	2	Q9ZKT5	helicobacte
15	7	3.2	710	2	O14407	neurospora
16	7	3.2	766	10	Q39177	arabidopsis
17	7	3.2	845	11	Q63370	rattus norv
18	7	3.2	988	2	Q9X537	corynebacte
19	7	3.2	990	2	O06512	enterobacte
20	7	3.2	1112	3	O93787	rhizopus ol
21	6	2.7	35	5	Q26372	tribolium c
22	6	2.7	35	12	O78063	human immun
23	6	2.7	44	12	O84304	human papil
24	6	2.7	44	12	O84305	human papil
25	6	2.7	44	12	O84309	human papil
26	6	2.7	44	12	O84315	human papil
27	6	2.7	44	12	Q84914	human papil
28	6	2.7	53	12	O97951	human immun
29	6	2.7	69	8	O29179	archaeoglob
30	6	2.7	71	8	Q35310	oryza sativ
31	6	2.7	79	12	Q04252	human papil
32	6	2.7	81	4	O43687	homo sapien
33	6	2.7	92	12	O80286	human immun
34	6	2.7	93	6	Q9ZBV2	streptomyce
35	6	2.7	94	6	Q47019	escherichia
36	6	2.7	112	11	Q9WU87	mus musculu
37	6	2.7	113	2	O33159	mycobacteri
38	6	2.7	117	10	O80978	arabidopsis
39	6	2.7	118	12	O56628	helicoverpa
40	6	2.7	118	12	O56631	tenebrio·mo
41	6	2.7	118	12	O56632	apis cerana
42	6	2.7	118	12	O56635	anticarsia
43	6	2.7	119	5	O01694	caenorhabdi
44	6	2.7	128	12	O81850	human papil
45	6	2.7	132	6	O77762	canis famil
46	6	2.7	132	12	O81852	human papil
47	6	2.7	137	6	Q9XS58	tursiops tr
48	6	2.7	138	10	P93856	skeletonema
49	6	2.7	139	5	Q19840	caenorhabdi
50	6	2.7	140	2	Q53856	spiroplasma
51	6	2.7	141	3	Q50881	myxococcus
52	6	2.7	142	5	O08687	saccharomyc
53	6	2.7	142	2	O00917	plasmodium
54	6	2.7	145	5	O00921	plasmodium
55	6	2.7	146	2	Q9X864	streptomyce
56	6	2.7	146	2	O52243	acinetobact
57	6	2.7	146	2	O52245	acinetobact
58	6	2.7	148	3	O43899	acinetobact
59	6	2.7	151	3	O07880	saccharomyc
60	6	2.7	153	9	Q26170	plasmodium
61	6	2.7	154	9	Q9XCI5	shigella fl
62	6	2.7	156	9	O21999	bacteriopha
63	6	2.7	171	9	O05745	saccharomyc
64	6	2.7	171	1	O55052	streptococc
65	6	2.7	172	4	Q03839	homo sapien
66	6	2.7	172	5	O00923	plasmodium
67	6	2.7	172	5	O00918	plasmodium
68	6	2.7	172	5	O00919	plasmodium
69	6	2.7	173	5	O00920	plasmodium
70	6	2.7	173	5	O00922	plasmodium
71	6	2.7	173	5	Q26185	plasmodium
72	6	2.7	180	5	O34479	bacillus su
73	6	2.7	180	9	P73586	synechocyst
74	6	2.7	182	2	O64175	bacteriopha
75	6	2.7	183	2	Q9X604	salmonella
76	6	2.7	183	3	O74882	schizosacch
77	6	2.7	183	3	O84877	human polio
78	6	2.7	183	12	Q66304	coxsackievi
79	6	2.7	183	12	Q66305	coxsackievi
80	6	2.7	183	12	Q66306	coxsackievi
81	6	2.7	183	12	Q66307	coxsackievi
82	6	2.7	183	12	Q66308	coxsackievi
83	6	2.7	183	12	Q66309	coxsackievi
84	6	2.7	183	12	Q66310	coxsackievi
85	6	2.7	183	12	Q66311	coxsackievi
86	6	2.7	183	12	Q66312	coxsackievi
87	6	2.7	183	12	Q66313	coxsackievi
88	6	2.7	183	12	Q66314	coxsackievi
89	6	2.7	183	12	Q66315	coxsackievi
90	6	2.7	183	12	Q66316	coxsackievi
91	6	2.7	183	12	Q66317	coxsackievi
92	6	2.7	183	12	Q66318	coxsackievi

Tue Jun 13 10:29:57 2000   us-09-252-691-7056.oligo.rspt

idx			acc	num	cnt	acc	name
93	2.7	6	Q66319	183	12	066319	coxsackievi
94	2.7	6	Q66320	183	12	066320	coxsackievi
95	2.7	6	Q90207	183	12	090207	coxsackievi
96	2.7	6	Q90208	183	12	090208	coxsackievi
97	2.7	6	Q92RQ9	183	12	092RQ9	coxsackievi
98	2.7	6	Q9zrq9	189	10	Q9zrq9	pseudotsuga
99	2.7	6	086520	189	2	086520	streptomyce
100	2.7	6	P93855	190	10	P93855	skeletonema
101	2.7	6	016565	194	5	016565	caenorhabdi
102	2.7	6	084301	199	12	084301	human papil
103	2.7	6	054128	209	2	054128	streptomyce
104	2.7	6	Q9X6F7	213	2	Q9x6f7	campylobact
105	2.7	6	Q9X6F5	213	2	Q9x6f5	campylobact
106	2.7	6	Q12143	216	3	Q12143	saccharomyc
107	2.7	6	P71487	221	2	P71487	methylomicr
108	2.7	6	Q9YMN7	223	12	Q9ymn7	lymantria d
109	2.7	6	Q42436	225	10	Q42436	brassica ra
110	2.7	6	Q50447	228	2	Q50447	mycobacteri
111	2.7	6	Q92506	228	2	Q92506	streptomyce
112	2.7	6	Q9X0E1	233	2	Q9x0e1	thermotoga
113	2.7	6	Q51240	235	2	Q51240	borrelia bu
114	2.7	6	Q9XCI4	235	2	Q9xci4	shigella fl
115	2.7	6	P74198	238	2	P74198	synechocyst
116	2.7	6	Q9WXX8	240	2	Q9wxx8	thermotoga
117	2.7	6	Q9XXH3	240	5	Q9xyh3	babesia big
118	2.7	6	086961	241	2	086961	thermotoga
119	2.7	6	P96892	242	2	P96892	mycobacteri
120	2.7	6	Q75832	245	12	Q75832	human immun
121	2.7	6	Q75834	246	12	Q75834	human immun
122	2.7	6	Q54363	250	2	Q54363	streptomyce
123	2.7	6	Q48792	251	2	Q48792	listeria mo
124	2.7	6	Q9AV7	253	1	Q9av7	aeropyrum p
125	2.7	6	Q50757	254	4	Q50757	borrelia bu
126	2.7	6	Q15466	256	2	Q15466	homo sapien
127	2.7	6	Q52617	257	2	Q52617	pseudomonas
128	2.7	6	Q69009	257	2	Q69009	pseudomonas
129	2.7	6	Q59648	257	2	Q59648	pseudomonas
130	2.7	6	P89923	257	12	P89923	bacteriopha
131	2.7	6	P71968	258	2	P71968	mycobacteri
132	2.7	6	P96163	258	2	P96163	vibrio furn
133	2.7	6	Q54172	261	2	Q54172	streptomyce
134	2.7	6	Q86975	262	12	Q86975	phocid herp
135	2.7	6	066158	265	2	066158	streptomyce
136	2.7	6	Q51400	266	2	Q51400	pseudomonas
137	2.7	6	005684	268	2	005684	mycobacteri
138	2.7	6	084991	269	2	084991	rhodococcus
139	2.7	6	Q9X9X6	269	2	Q9x9x6	streptomyce
140	2.7	6	Q93721	270	5	Q93721	caenorhabdi
141	2.7	6	Q9X2I3	271	2	Q9x2i3	thermotoga
142	2.7	6	Q49993	273	2	Q49993	mycobacteri
143	2.7	6	P71671	274	2	P71671	mycobacteri
144	2.7	6	065599	276	10	065599	arabidopsis
145	2.7	6	Q27045	277	2	Q27045	theileria p
146	2.7	6	059568	281	1	059568	pyrococcus
147	2.7	6	087598	281	5	087598	streptomyce
148	2.7	6	Q9WYB5	281	2	Q9wyb5	thermotoga
149	2.7	6	Q9ZWP5	285	10	Q9zwp5	robinia pse
150	2.7	6	058132	288	2	058132	pyrococcus
151	2.7	6	Q9X7T0	294	2	Q9x7t0	streptomyce
152	2.7	6	Q9YH87	294	13	Q9yh87	oryzias lat
153	2.7	6	Q9X9Q1	296	2	Q9x9q1	streptomyce
154	2.7	6	041026	298	12	041026	paramecium
155	2.7	6	066788	300	2	066788	aquifex aeo
156	2.7	6	Q92XD7	300	9	Q9xd7	bacteriopha
157	2.7	6	018331	301	2	018331	hemicentrot
158	2.7	6	000801	304	5	000801	plasmodium
159	2.7	6	Q92BV0	307	2	Q9bv0	streptomyce
160	2.7	6	027397	312	1	027397	methanobact
161	2.7	6	000796	312	5	000796	plasmodium
162	2.7	6	000797	312	5	000797	plasmodium
163	2.7	6	088034	313	2	088034	streptomyce
164	2.7	6	000794	315	5	000794	plasmodium
165	2.7	6	000798	315	5	000798	plasmodium

name	acc	acc	cnt	num			idx
plasmodium	000799	000799	5	315	2.7	6	166
plasmodium	000802	000802	5	315	2.7	6	167
plasmodium	000803	000803	5	315	2.7	6	168
synechococc	P96034	P96034	5	317	2.7	6	169
bacillus su	034966	034966	2	319	2.7	6	170
mycobacteri	Q59553	Q59553	2	321	2.7	6	171
mycobacteri	P96240	P96240	2	322	2.7	6	172
homo sapien	075370	075370	4	324	2.7	6	173
caenorhabdi	Q9XXP7	Q9XXP7	5	328	2.7	6	174
yersinia pe	068786	068786	5	328	2.7	6	175
caenorhabdi	045299	045299	5	329	2.7	6	176
toxoplasma	P90613	P90613	5	331	2.7	6	177
caenorhabdi	Q17810	Q17810	5	333	2.7	6	178
plasmodium	000795	000795	5	336	2.7	6	179
actinosynne	069207	069207	3	338	2.7	6	180
schizosacch	043083	043083	3	338	2.7	6	181
plasmodium	000791	000791	5	338	2.7	6	182
plasmodium	000793	000793	5	339	2.7	6	183
plasmodium	096270	096270	5	339	2.7	6	184
plasmodium	000792	000792	5	339	2.7	6	185
plasmodium	000800	000800	5	342	2.7	6	186
treponema p	083088	083088	2	344	2.7	6	187
caenorhabdi	045427	045427	5	345	2.7	6	188
haemophilus	086241	086241	2	351	2.7	6	189
pseudomonas	005599	005599	5	351	2.7	6	190
spinacia ol	P93547	P93547	10	353	2.7	6	191
rice dwarf	Q85447	Q85447	12	355	2.7	6	192
rhesus cyto	Q9YUG8	Q9YUG8	12	356	2.7	6	193
caenorhabdi	Q9XTB2	Q9XTB2	5	360	2.7	6	194
mycobacteri	053411	053411	2	360	2.7	6	195
treponema d	087547	087547	2	361	2.7	6	196
agrobacteri	Q44198	Q44198	2	363	2.7	6	197
hepatitis c	Q81549	Q81549	12	364	2.7	6	198
hepatitis c	Q81551	Q81551	12	364	2.7	6	199
hepatitis c	Q81716	Q81716	12	365	2.7	6	200
nelson bay	072468	072468	12	367	2.7	6	201
lactobacill	053047	053047	2	368	2.7	6	202
homo sapien	Q9Y360	Q9Y360	4	370	2.7	6	203
chlamydia p	Q927K1	Q927K1	2	371	2.7	6	204
herpes simp	P89439	P89439	12	372	2.7	6	205
vibrio furn	059445	059445	1	377	2.7	6	206
arabidopsis	048680	048680	10	377	2.7	6	207
bos taurus	Q9XS61	Q9XS61	6	382	2.7	6	208
aeropyrum p	Q9YFP0	Q9YFP0	1	383	2.7	6	209
thermotoga	Q9WZ93	Q9WZ93	2	383	2.7	6	210
streptomyce	Q9X829	Q9X829	2	386	2.7	6	211
xenopus lae	Q91839	Q91839	13	386	2.7	6	212
aeropyrum p	Q9Z8Q9	Q9Z8Q9	1	390	2.7	6	213
actinobacil	Q9ZID8	Q9ZID8	2	390	2.7	6	214
fugu rubrip	Q9W6U5	Q9W6U5	13	390	2.7	6	215
amycolatops	052822	052822	2	391	2.7	6	216
synechocyst	P72824	P72824	2	391	2.7	6	217
caenorhabdi	017327	017327	5	391	2.7	6	218
aeropyrum p	Q9YBB6	Q9YBB6	1	393	2.7	6	219
physarum po	003486	003486	5	402	2.7	6	220
fugu rubrip	Q9W5Z9	Q9W5Z9	13	402	2.7	6	221
bordetella	Q45380	Q45380	2	403	2.7	6	222
streptomyce	069941	069941	2	403	2.7	6	223
amycolatops	052823	052823	2	406	2.7	6	224
synechocyst	031557	031557	2	407	2.7	6	225
bacillus su	059721	059721	3	407	2.7	6	226
arabidopsis	082257	082257	10	407	2.7	6	227
arabidopsis	034765	034765	2	408	2.7	6	228
bacillus su	Q9XF6	Q9XF6	5	409	2.7	6	229
chaetoptery	Q65354	Q65354	12	412	2.7	6	230
autographa	062419	062419	2	413	2.7	6	231
caenorhabdi	Q61395	Q61395	11	414	2.7	6	232
mus musculu	Q05732	Q05732	11	414	2.7	6	233
mus musculu	025667	025667	2	416	2.7	6	234
helicobacte	Q92M22	Q92M22	2	416	2.7	6	235
helicobacte	087999	087999	2	419	2.7	6	236
bordetella	P97610	P97610	11	421	2.7	6	237
rattus norv	054505	054505	2	422	2.7	6	238
salmonella							

#			Accession		Description	Grp
239	2.7	6	Q9WYM6	425	Q9WYM6 thermotoga	2
240	2.7	6	O45500	433	O45500 caenorhabdi	5
241	2.7	6	Q97964	436	Q97964 aeropyrum p	5
242	2.7	6	Q9XGG0	436	Q9xgg0 aeropyrum p	10
243	2.7	6	Q9XGF9	436	Q9xgf9 selaginella	10
244	2.7	6	O35249	437	O35249 rattus norv	11
245	2.7	6	O81783	438	O81783 arabidopsis	10
246	2.7	6	Q9WV78	438	Q9wv78 rattus norv	11
247	2.7	6	Q14314	439	Q14314 homo sapien	4
248	2.7	6	Q9Y122	439	Q9y122 drosophila	2
249	2.7	6	O27792	444	O27792 methanobact	1
250	2.7	6	Q55937	445	Q55937 synechocyst	2
251	2.7	6	P76097	447	P76097 escherichia	2
252	2.7	6	Q62162	447	Q62162 mus musculu	11
253	2.7	6	Q924F7	448	Q924f7 escherichia	2
254	2.7	6	O50924	450	O50924 borrelia bu	2
255	2.7	6	O17120	456	O17120 caenorhabdi	5
256	2.7	6	Q9YBZ2	461	Q9ybz2 aeropyrum p	1
257	2.7	6	Q9XBP5	465	Q9xbp5 myxococcus	2
258	2.7	6	Q56291	466	Q56291 thiobacillu	5
259	2.7	6	O70527	467	O70527 cavia porce	11
260	2.7	6	O77466	469	O77466 drosophila	5
261	2.7	6	O53446	474	O53446 mycobacteri	2
262	2.7	6	Q9X789	474	Q9x789 mycobacteri	2
263	2.7	6	Q9XWI7	485	Q9xwi7 caenorhabdi	5
264	2.7	6	O83778	490	O83778 treponema p	2
265	2.7	6	Q9XA17	490	Q9xa17 streptomyce	2
266	2.7	6	O76191	497	O76191 dirofilaria	5
267	2.7	6	O54116	498	O54116 streptomyce	2
268	2.7	6	Q9WBS1	500	Q9wbs1 human papil	12
269	2.7	6	Q9WBS0	500	Q9wbs0 human papil	12
270	2.7	6	Q9WBR9	500	Q9wbr9 human papil	12
271	2.7	6	Q9WBR8	500	Q9wbr8 human papil	12
272	2.7	6	Q9WBR7	500	Q9wbr7 human papil	12
273	2.7	6	Q9WBR6	500	Q9wbr6 human papil	12
274	2.7	6	Q9WBR5	500	Q9wbr5 human papil	12
275	2.7	6	Q9W9C6	500	Q9w9c6 human papil	12
276	2.7	6	O50157	501	O50157 streptococc	2
277	2.7	6	O37391	501	O37391 common chim	12
278	2.7	6	O82010	503	O82010 human papil	12
279	2.7	6	O81958	503	O81958 human papil	12
280	2.7	6	Q9WNM4	503	Q9wnm4 human papil	12
281	2.7	6	O81961	504	O81961 human papil	12
282	2.7	6	O98923	504	O98923 gallus gall	13
283	2.7	6	Q9YXD1	505	Q9yxd1 human papil	12
284	2.7	6	Q9XEG4	506	Q9xeg4 hordeum vul	10
285	2.7	6	O82004	506	O82004 human papil	12
286	2.7	6	P93182	507	P93182 hordeum vul	10
287	2.7	6	Q9XEG5	507	Q9xeg5 hordeum vul	10
288	2.7	6	Q81971	508	Q81971 human papil	12
289	2.7	6	P93183	509	P93183 hordeum vul	10
290	2.7	6	Q9Z1K8	510	Q9z1k8 mus musculu	11
291	2.7	6	P89427	510	P89427 human papil	12
292	2.7	6	Q924W7	511	Q924w7 streptomyce	2
293	2.7	6	O81709	511	O81709 arabidopsis	10
294	2.7	6	P94684	512	P94684 chlamydia t	2
295	2.7	6	Q53769	512	Q53769 streptomyce	2
296	2.7	6	O01870	512	O01870 caenorhabdi	5
297	2.7	6	Q9YG94	514	Q9yg94 aeropyrum p	1
298	2.7	6	Q9XAJ9	514	Q9xaj9 streptomyce	2
299	2.7	6	P91036	514	P91036 caenorhabdi	5
300	2.7	6	P91137	517	P91137 caenorhabdi	5
301	2.7	6	O23240	524	O23240 arabidopsis	10
302	2.7	6	Q9Y2K4	525	Q9y2k4 homo sapien	4
303	2.7	6	Q928Y6	526	Q928y6 chlamydia p	2
304	2.7	6	O90730	529	O90730 human papil	12
305	2.7	6	Q92205	532	Q92205 botrytis ci	3
306	2.7	6	Q53903	533	Q53903 streptomyce	2
307	2.7	6	O82003	534	O82003 human papil	12
308	2.7	6	Q9Y4Y5	536	Q9y4y5 homo sapien	4
309	2.7	6	Q25332	537	Q25332 leishmania	5
310	2.7	6	Q24281	540	Q24281 drosophila	5
311	2.7	6	Q40283	541	Q40283 manihot esc	10

Description	Accession	Grp	Pos			#
O29248 archaeoglob	O29248	1	547	2.7	6	312
O45812 caenorhabdi	O45812	5	552	2.7	6	313
O26642 sarcophaga	O26642	5	553	2.7	6	314
O56285 human herpe	O56285	12	554	2.7	6	315
O34105 salmonella	O34105	2	561	2.7	6	316
O30916 salmonella	O30916	5	563	2.7	6	317
O27937 loligo forb	O27937	5	564	2.7	6	318
P87142 schizosacch	P87142	3	566	2.7	6	319
O28188 archaeoglob	O28188	5	569	2.7	6	320
O17654 caenorhabdi	O17654	5	573	2.7	6	321
Q98921 gallus gall	Q98921	13	584	2.7	6	322
Q90989 gallus gall	Q90989	13	584	2.7	6	323
O53417 mycobacteri	O53417	2	587	2.7	6	324
O23340 arabidopsis	O23340	10	587	2.7	6	325
Q01720 homo sapien	Q01720	5	591	2.7	6	326
O62306 caenorhabdi	O62306	5	596	2.7	6	327
Q92DQ1 rickettsia	Q92DQ1	2	600	2.7	6	328
P90858 caenorhabdi	P90858	5	611	2.7	6	329
Q01203 melampsora	Q01203	3	614	2.7	6	330
O96437 eimeria ten	O96437	5	617	2.7	6	331
O00926 trypanosoma	O00926	2	622	2.7	6	332
O17082 caenorhabdi	O17082	5	622	2.7	6	333
Q98922 gallus gall	Q98922	13	626	2.7	6	334
Q90880 gallus gall	Q90880	13	626	2.7	6	335
O26182 plasmodium	O26182	5	636	2.7	6	336
O67583 aquifex aeo	O67583	2	638	2.7	6	337
Q07528 saccharomyc	Q07528	2	640	2.7	6	338
Q84876 human polio	Q84876	12	643	2.7	6	339
O52071 lactobacill	O52071	2	647	2.7	6	340
Q96714 chlorella v	Q96714	12	648	2.7	6	341
O41043 paramecium	O41043	12	649	2.7	6	342
Q27150 oxytricha n	Q27150	5	651	2.7	6	343
O28469 archaeoglob	O28469	1	663	2.7	6	344
O45018 borrelia bu	O45018	2	672	2.7	6	345
O45041 borrelia bu	O45041	2	672	2.7	6	346
P90755 caenorhabdi	P90755	2	674	2.7	6	347
O83759 treponema p	O83759	2	679	2.7	6	348
O00838 leishmania	O00838	5	683	2.7	6	349
Q55846 synechocyst	Q55846	5	684	2.7	6	350
O62527 drosophila	O62527	5	687	2.7	6	351
O45101 caenorhabdi	O45101	5	692	2.7	6	352
Q45016 borrelia bu	Q45016	2	693	2.7	6	353
Q44937 borrelia bu	Q44937	2	693	2.7	6	354
Q45043 borrelia bu	Q45043	2	693	2.7	6	355
Q45163 borrelia ga	Q45163	2	695	2.7	6	356
O52753 ruminococcu	O52753	2	695	2.7	6	357
O26996 methanobact	O26996	1	698	2.7	6	358
O44850 caenorhabdi	O44850	5	698	2.7	6	359
O00839 leishmania	O00839	5	707	2.7	6	360
Q83072 leishmania	Q83072	12	714	2.7	6	361
Q48553 leuconostoc	Q48553	5	717	2.7	6	362
Q9Y4W2 homo sapien	Q9Y4W2	4	717	2.7	6	363
O94439 schizosacch	O94439	3	728	2.7	6	364
O70585 mus musculu	O70585	11	730	2.7	6	365
Q24557 drosophila	Q24557	5	732	2.7	6	366
Q59490 lactobacill	Q59490	2	738	2.7	6	367
O60327 homo sapien	O60327	5	738	2.7	6	368
O93967 amanita mus	O93967	3	740	2.7	6	369
O88408 mus musculu	O88408	11	746	2.7	6	370
O84875 human polio	O84875	12	752	2.7	6	371
Q9X5P3 pseudomonas	Q9X5P3	2	756	2.7	6	372
Q85082 human polio	Q85082	12	757	2.7	6	373
O48795 arabidopsis	O48795	10	760	2.7	6	374
O33466 pseudomonas	O33466	2	766	2.7	6	375
Q9WW20 pseudomonas	Q9WW20	2	766	2.7	6	376
O85783 myxococcus	O85783		770	2.7	6	377
Q96325 arabidopsis	Q96325	10	773	2.7	6	378
O64990 arabidopsis	O64990	10	775	2.7	6	379
P92941 arabidopsis	P92941	10	775	2.7	6	380
O98167 lapine rota	O98167	12	776	2.7	6	381
O98168 lapine rota	O98168	12	776	2.7	6	382
P92942 arabidopsis	P92942	10	780	2.7	6	383
O65383 arabidopsis	O65383	10	783	2.7	6	384

Left block (rows 385–457)

No.	v1	v2	pos	n	accession	organism
385	2.7	6	786	2	O32044	bacillus su
386	2.7	6	786	10	Q9XF71	nicotiana t
387	2.7	6	790	5	O01264	caenorhabdi
388	2.7	6	802	5	Q19764	caenorhabdi
389	2.7	6	805	3	O94133	piromyces s
390	2.7	6	814	3	Q05958	saccharomyc
391	2.7	6	825	2	Q48760	listeria mo
392	2.7	6	835	2	Q9X0A6	thermotoga
393	2.7	6	840	4	O14587	homo sapien
394	2.7	6	841	2	Q92J88	helicobacte
395	2.7	6	842	5	O25950	helicobacte
396	2.7	6	864	10	Q42847	hordeum vul
397	2.7	6	873	2	O52598	bradyrhizob
398	2.7	6	875	10	Q42739	flaveria tr
399	2.7	6	877	10	Q42497	chlorella v
400	2.7	6	879	3	Q07807	saccharomyc
401	2.7	6	879	11	Q9WV91	mus musculu
402	2.7	6	884	10	O23928	eleocharis
403	2.7	6	887	10	O82032	oryza sativ
404	2.7	6	888	5	Q25336	leishmania
405	2.7	6	890	11	Q921A0	cavia porce
406	2.7	6	896	11	Q64146	rattus norv
407	2.7	6	907	10	Q42738	flaveria tr
408	2.7	6	912	12	O90278	chimpanzee
409	2.7	6	913	4	Q9Y2K6	homo sapien
410	2.7	6	919	12	Q9WPP0	chimpanzee
411	2.7	6	921	10	Q24253	drosophila
412	2.7	6	947	5	O23927	eleocharis
413	2.7	6	947	10	O24612	oryza sativ
414	2.7	6	950	10	Q41847	zea mays (m
415	2.7	6	960	12	Q66678	equine herp
416	2.7	6	980	10	O23404	arabidopsis
417	2.7	6	990	12	Q66682	equine herp
418	2.7	6	990	4	Q15206	homo sapien
419	2.7	6	997	10	Q40699	oryza sativ
420	2.7	6	1003	12	Q92411	bacillus su
421	2.7	6	1003	1	O58064	pyrococcus
422	2.7	6	1010	5	Q9XXR9	caenorhabdi
423	2.7	6	1014	13	Q9YHU7	xenopus lae
424	2.7	6	1038	5	O61702	limulus pol
425	2.7	6	1049	2	O05924	neisseria m
426	2.7	6	1067	2	O45572	caenorhabdi
427	2.7	6	1079	13	Q9YHU6	xenopus lae
428	2.7	6	1103	3	Q59599	n carbamoyl
429	2.7	6	1108	12	O41928	murine herp
430	2.7	6	1139	2	Q9Y7P5	schizosacch
431	2.7	6	1142	12	O54073	synechococc
432	2.7	6	1194	13	P89462	herpes simp
433	2.7	6	1218	5	Q9W737	gallus gall
434	2.7	6	1232	5	O05331	homo sapien
435	2.7	6	1234	5	P94692	desulfovibr
436	2.7	6	1240	13	O01505	caenorhabdi
437	2.7	6	1243	11	P79773	gallus gall
438	2.7	6	1244	4	O35954	mus musculu
439	2.7	6	1270	5	O00562	homo sapien
440	2.7	6	1288	5	Q19736	caenorhabdi
441	2.7	6	1329	10	O13682	schizosacch
442	2.7	6	1330	12	Q9ZUK1	arabidopsis
443	2.7	6	1344	11	Q04534	pseudorabie
444	2.7	6	1350	5	Q60697	mus musculu
445	2.7	6	1364	4	Q27571	drosophila
446	2.7	6	1381	11	O75981	homo sapien
447	2.7	6	1382	4	P97846	rattus norv
448	2.7	6	1384	4	Q14152	homo sapien
449	2.7	6	1385	5	P78357	homo sapien
450	2.7	6	1385	5	Q26388	drosophila
451	2.7	6	1387	11	O54991	mus musculu
452	2.7	6	1389	13	O57512	anas platyr
453	2.7	6	1389	5	Q24591	drosophila
454	2.7	6	1395	13	Q90240	anas sp. (d
455	2.7	6	1413	8	Q35059	marchantia
456	2.7	6	1413	13	O75696	homo sapien
457	2.7	6	1475	10	Q9XEP3	sorghum bic

Right block (rows 458–530)

No.	v1	v2	pos	n	accession	organism
458	2.7	6	1494	5	Q94885	drosophila
459	2.7	6	1510	4	Q9Y4G3	homo sapien
460	2.7	6	1516	4	P78354	homo sapien
461	2.7	6	1524	5	O15830	trypanosoma
462	2.7	6	1532	4	P78544	homo sapien
463	2.7	6	1549	4	O60706	homo sapien
464	2.7	6	1549	4	O60707	homo sapien
465	2.7	6	1562	5	Q92GI3	streptomyce
466	2.7	6	1592	5	Q19386	caenorhabdi
467	2.7	6	1684	3	O00310	homo sapien
468	2.7	6	1710	3	O74136	candida alb
469	2.7	6	1724	3	O25950	helicobacte
470	2.7	6	1726	5	O93927	cryptococcu
471	2.7	6	1741	5	O02569	plasmodium
472	2.7	6	1751	5	O00498	homo sapien
473	2.7	6	1767	5	O26194	plasmodium
474	2.7	6	1781	5	O24495	drosophila
475	2.7	6	1820	5	Q99970	homo sapien
476	2.7	6	1857	5	Q9XTF0	caenorhabdi
477	2.7	6	1862	5	O95153	homo sapien
478	2.7	6	2051	5	O20090	caenorhabdi
479	2.7	6	2080	5	O44328	hirudo medi
480	2.7	6	2091	3	O75923	homo sapien
481	2.7	6	2118	3	P78616	emericella
482	2.7	6	2206	6	O76904	drosophila
483	2.7	6	2207	12	Q84792	human polio
484	2.7	6	2221	12	Q98595	human polio
485	2.7	6	2251	12	O84865	human polio
486	2.7	6	2342	12	O11388	porcine rub
487	2.7	6	2433	12	Q65980	cherry capi
488	2.7	6	2514	12	O91464	aichi virus
489	2.7	6	2533	5	Q9Y061	caenorhabdi
490	2.7	6	2533	5	O27183	paramecium
491	2.7	6	2619	5	P90589	paramecium
492	2.7	6	2628	5	O22258	caenorhabdi
493	2.7	6	2705	3	O06179	saccharomyc
494	2.7	6	2731	13	Q9W6V6	gallus gall
495	2.7	6	2731	2	Q9Z3T9	pseudomonas
496	2.7	6	2809	11	Q9WTS4	mus musculu
497	2.7	6	2894	1	O61230	lytechinus
498	2.7	6	3011	12	Q58791	methanococc
499	2.7	6	3011	12	O36579	hepatitis c
500	2.7	6	3011	12	O36608	hepatitis c
501	2.7	6	3011	12	O36609	hepatitis c
502	2.7	6	3052	12	O36610	hepatitis c
503	2.7	6	3052	12	Q03463	hepatitis c
504	2.7	6	3124	12	O82933	johnsongras
505	2.7	6	3161	2	O66237	citrus tris
506	2.7	6	3163	2	O54511	yersinia en
507	2.7	6	3413	2	O92373	yersinia pe
508	2.7	6	3851	4	O54593	amycolatops
509	2.7	6	4919	5	O43161	homo sapien
510	2.7	6	6658	—	Q9ZHL0	haemophilus
511	2.3	5	9	4	O76281	drosophila
512	2.3	5	15	5	O95574	homo sapien
513	2.3	5	16	4	P78533	homo sapien
514	2.3	5	19	11	Q9ZOZ2	mus musculu
515	2.3	5	19	8	O63058	lathraea cl
516	2.3	5	21	19	Q82392	human t-cel
517	2.3	5	21	21	O82410	human t-cel
518	2.3	5	21	21	O82411	human t-cel
519	2.3	5	21	21	O82412	human t-cel
520	2.3	5	26	10	O40536	nicotiana t
521	2.3	5	32	11	O64085	mus musculu
522	2.3	5	33	12	O73052	hepatitis c
523	2.3	5	33	12	O73053	hepatitis c
524	2.3	5	34	5	O61261	acyrthosiph
525	2.3	5	34	12	O71809	human immun
526	2.3	5	34	12	O71813	human immun
527	2.3	5	34	12	O71814	human immun
528	2.3	5	34	12	O79463	human immun
529	2.3	5	34	12	O79493	human immun
530	2.3	5	34	12	O91354	human immun
530	2.3	5	34	12	O91355	human immun

#	ID	Description	ID				
531	091356	human immun	091356	12	34	2.3	5
532	091357	human immun	091357	12	34	2.3	5
533	091358	human immun	091358	12	34	2.3	5
534	091359	human immun	091359	12	34	2.3	5
535	091360	human immun	091360	12	34	2.3	5
536	091361	human immun	091361	12	34	2.3	5
537	091365	human immun	091365	12	34	2.3	5
538	091367	human immun	091367	12	34	2.3	5
539	091368	human immun	091368	8	34	2.3	5
540	Q32706	nicotiana t	Q32706	10	35	2.3	5
541	Q39519	nicotiana t	Q39519	12	35	2.3	5
542	Q76251	human immun	Q76251	12	35	2.3	5
543	Q76374	human immun	Q76374	12	35	2.3	5
544	Q97629	human immun	Q97629	12	35	2.3	5
545	Q97644	human immun	Q97644	12	35	2.3	5
546	040580	human immun	040580	12	35	2.3	5
547	040586	human immun	040586	12	35	2.3	5
548	Q70414	human immun	Q70414	12	35	2.3	5
549	Q70427	human immun	Q70427	12	35	2.3	5
550	Q76296	human immun	Q76296	12	35	2.3	5
551	Q76369	human immun	Q76369	12	35	2.3	5
552	Q78035	human immun	Q78035	12	35	2.3	5
553	Q78039	human immun	Q78039	12	35	2.3	5
554	Q78043	human immun	Q78043	12	35	2.3	5
555	Q78055	human immun	Q78055	12	35	2.3	5
556	Q78059	human immun	Q78059	12	35	2.3	5
557	Q78094	human immun	Q78094	12	35	2.3	5
558	Q78102	human immun	Q78102	12	35	2.3	5
559	Q78107	human immun	Q78107	12	35	2.3	5
560	Q78110	human immun	Q78110	12	35	2.3	5
561	Q78198	human immun	Q78198	12	35	2.3	5
562	Q78201	human immun	Q78201	12	35	2.3	5
563	Q78212	human immun	Q78212	12	35	2.3	5
564	Q80012	human immun	Q80012	12	35	2.3	5
565	Q80465	human immun	Q80465	12	35	2.3	5
566	Q80476	human immun	Q80476	12	35	2.3	5
567	Q80478	human immun	Q80478	12	35	2.3	5
568	Q80519	human immun	Q80519	12	35	2.3	5
569	Q80590	human immun	Q80590	12	35	2.3	5
570	Q80592	human immun	Q80592	12	35	2.3	5
571	Q80603	human immun	Q80603	12	35	2.3	5
572	Q71189	human immun	Q71189	12	35	2.3	5
573	Q71190	human immun	Q71190	12	35	2.3	5
574	Q71191	human immun	Q71191	12	35	2.3	5
575	Q77996	human immun	Q77996	12	35	2.3	5
576	Q78026	human immun	Q78026	12	35	2.3	5
577	Q78046	human immun	Q78046	12	35	2.3	5
578	Q78048	human immun	Q78048	12	35	2.3	5
579	Q78051	human immun	Q78051	12	35	2.3	5
580	Q78056	human immun	Q78056	12	35	2.3	5
581	Q78065	human immun	Q78065	12	35	2.3	5
582	Q78067	human immun	Q78067	12	35	2.3	5
583	Q78071	human immun	Q78071	12	35	2.3	5
584	Q78072	human immun	Q78072	12	35	2.3	5
585	Q78075	human immun	Q78075	12	35	2.3	5
586	Q78079	human immun	Q78079	12	35	2.3	5
587	Q78081	human immun	Q78081	12	35	2.3	5
588	Q78082	human immun	Q78082	12	35	2.3	5
589	Q78083	human immun	Q78083	12	35	2.3	5
590	Q78084	human immun	Q78084	12	35	2.3	5
591	Q78087	human immun	Q78087	12	35	2.3	5
592	Q78091	human immun	Q78091	12	35	2.3	5
593	Q78093	human immun	Q78093	12	35	2.3	5
594	Q78096	human immun	Q78096	12	35	2.3	5
595	Q78097	human immun	Q78097	12	35	2.3	5
596	Q78103	human immun	Q78103	12	35	2.3	5
597	Q78104	human immun	Q78104	12	35	2.3	5
598	Q78105	human immun	Q78105	12	35	2.3	5
599	Q78106	human immun	Q78106	12	35	2.3	5
600	Q78111	human immun	Q78111	12	35	2.3	5
601	Q78122	human immun	Q78122	12	35	2.3	5
602	Q78136	human immun	Q78136	12	35	2.3	5
603	Q78139	human immun	Q78139	12	35	2.3	5
604	Q78145	human immun	Q78145	12	35	2.3	5
605	Q78175	human immun	Q78175	12	35	2.3	5
606	Q78186	human immun	Q78186	12	35	2.3	5
607	Q78206	human immun	Q78206	12	35	2.3	5
608	Q78213	human immun	Q78213	12	35	2.3	5
609	Q78215	human immun	Q78215	12	35	2.3	5
610	Q79484	human immun	Q79484	12	35	2.3	5
611	Q79500	human immun	Q79500	12	35	2.3	5
612	Q80477	human immun	Q80477	12	35	2.3	5
613	Q80591	human immun	Q80591	12	35	2.3	5
614	Q70780	human immun	Q70780	12	35	2.3	5
615	Q70781	human immun	Q70781	12	35	2.3	5
616	Q70782	human immun	Q70782	12	35	2.3	5
617	Q70783	human immun	Q70783	12	35	2.3	5
618	Q70810	human immun	Q70810	12	35	2.3	5
619	Q70811	human immun	Q70811	12	35	2.3	5
620	Q78492	human immun	Q78492	12	35	2.3	5
621	Q78493	human immun	Q78493	12	35	2.3	5
622	010869	human immun	010869	12	35	2.3	5
623	090503	human immun	090503	12	35	2.3	5
624	090513	human immun	090513	12	35	2.3	5
625	090525	human immun	090525	12	35	2.3	5
626	091349	human immun	091349	12	35	2.3	5
627	091350	human immun	091350	12	35	2.3	5
628	091351	human immun	091351	12	35	2.3	5
629	091352	human immun	091352	12	35	2.3	5
630	091353	human immun	091353	12	35	2.3	5
631	091430	human immun	091430	12	35	2.3	5
632	091433	human immun	091433	12	35	2.3	5
633	091434	human immun	091434	12	35	2.3	5
634	091435	human immun	091435	12	35	2.3	5
635	091436	human immun	091436	12	35	2.3	5
636	091437	human immun	091437	12	35	2.3	5
637	091439	human immun	091439	12	35	2.3	5
638	091440	human immun	091440	12	35	2.3	5
639	091441	human immun	091441	12	35	2.3	5
640	091442	human immun	091442	12	35	2.3	5
641	091443	human immun	091443	12	35	2.3	5
642	091444	human immun	091444	12	35	2.3	5
643	091445	human immun	091445	12	35	2.3	5
644	091446	human immun	091446	12	35	2.3	5
645	091447	human immun	091447	12	35	2.3	5
646	Q76276	human immun	Q76276	12	35	2.3	5
647	Q77931	human immun	Q77931	12	35	2.3	5
648	Q79420	human immun	Q79420	12	35	2.3	5
649	Q76295	human immun	Q76295	12	35	2.3	5
650	Q70232	human immun	Q70232	12	35	2.3	5
651	Q9WCX4	human immun	Q9WCX4	12	35	2.3	5
652	Q9WCW6	human immun	Q9WCW6	12	35	2.3	5
653	Q9WCW5	human immun	Q9WCW5	12	35	2.3	5
654	Q9W8V8	human immun	Q9W8V8	12	35	2.3	5
655	010487	human immun	010487	12	37	2.3	5
656	024580	zea mays (m	024580	10	38	2.3	5
657	Q72317	human immun	Q72317	12	42	2.3	5
658	072845	human immun	072845	12	43	2.3	5
659	072846	human immun	072846	12	43	2.3	5
660	072847	human immun	072847	12	43	2.3	5
661	072854	human immun	072854	12	43	2.3	5
662	072859	human immun	072859	12	43	2.3	5
663	072869	human immun	072869	12	43	2.3	5
664	072870	human immun	072870	12	43	2.3	5
665	P97550	rattus norv	P97550	11	44	2.3	5
666	Q84307	human papil	Q84307	12	44	2.3	5
667	Q84308	human papil	Q84308	12	44	2.3	5
668	092307	human respi	092307	12	44	2.3	5
669	092315	human respi	092315	12	44	2.3	5
670	092318	human respi	092318	12	44	2.3	5
671	Q9WJB0	human immun	Q9WJB0	12	44	2.3	5
672	Q9WA44	human immun	Q9WA44	12	44	2.3	5
673	Q62307	mus musculu	Q62307	11	45	2.3	5
674	071191	grapevine l	071191	12	45	2.3	5
675	Q9WCW2	human immun	Q9WCW2	12	45	2.3	5
676	O50275	rhodococcus	O50275	2	46	2.3	5

#					Accession	Description
677	5	2.3	46	10	Q040923	pseudotsuga
678	5	2.3	46	12	Q072529	human immun
679	5	2.3	46	12	Q082401	human t-cel
680	5	2.3	46	12	Q082402	human t-cel
681	5	2.3	46	12	Q082403	human t-cel
682	5	2.3	46	12	Q082404	human t-cel
683	5	2.3	46	12	Q082407	human t-cel
684	5	2.3	48	11	Q063885	mus musculu
685	5	2.3	48	12	Q091284	human immun
686	5	2.3	49	12	Q50263	phytoplasma
687	5	2.3	49	12	Q9YL02	human immun
688	5	2.3	50	3	P89114	saccharomyc
689	5	2.3	50	12	Q9WMM9	human polio
690	5	2.3	50	12	Q9WMM8	human polio
691	5	2.3	51	11	O88660	rattus norv
692	5	2.3	51	12	O37475	human immun
693	5	2.3	52	5	O97152	drosophila
694	5	2.3	53	2	Q57170	clostridium
695	5	2.3	53	4	Q9Y6W9	homo sapien
696	5	2.3	53	10	Q41894	zea mays (m
697	5	2.3	53	12	Q69812	human immun
698	5	2.3	53	12	Q69828	human immun
699	5	2.3	53	12	Q76523	human immun
700	5	2.3	53	12	Q76546	human immun
701	5	2.3	53	12	Q97824	human immun
702	5	2.3	53	12	Q97826	human immun
703	5	2.3	53	12	Q97831	human immun
704	5	2.3	53	12	Q97835	human immun
705	5	2.3	53	12	Q97842	human immun
706	5	2.3	53	12	Q97843	human immun
707	5	2.3	53	12	Q97844	human immun
708	5	2.3	53	12	Q97949	human immun
709	5	2.3	53	12	Q97950	human immun
710	5	2.3	53	12	Q97954	human immun
711	5	2.3	53	12	Q97957	human immun
712	5	2.3	53	12	Q97958	human immun
713	5	2.3	53	12	Q97962	human immun
714	5	2.3	53	12	Q97965	human immun
715	5	2.3	53	12	Q97967	human immun
716	5	2.3	53	12	Q97968	human immun
717	5	2.3	53	12	Q97973	human immun
718	5	2.3	53	12	Q97820	human immun
719	5	2.3	53	12	Q97821	human immun
720	5	2.3	53	12	Q97822	human immun
721	5	2.3	53	12	Q97823	human immun
722	5	2.3	53	12	Q97825	human immun
723	5	2.3	53	12	Q97828	human immun
724	5	2.3	53	12	Q97829	human immun
725	5	2.3	53	12	Q97833	human immun
726	5	2.3	53	12	Q97836	human immun
727	5	2.3	53	12	Q97837	human immun
728	5	2.3	53	12	Q97840	human immun
729	5	2.3	53	12	Q97841	human immun
730	5	2.3	53	12	Q97952	human immun
731	5	2.3	53	12	Q97953	human immun
732	5	2.3	53	12	Q97955	human immun
733	5	2.3	53	12	Q97956	human immun
734	5	2.3	53	12	Q97960	human immun
735	5	2.3	53	12	Q97961	human immun
736	5	2.3	53	12	Q97963	human immun
737	5	2.3	53	12	Q97964	human immun
738	5	2.3	53	12	Q97969	human immun
739	5	2.3	53	12	Q97970	human immun
740	5	2.3	53	12	Q97971	human immun
741	5	2.3	53	12	Q97972	human immun
742	5	2.3	53	12	O37554	hepatitis c
743	5	2.3	53	12	O37556	hepatitis c
744	5	2.3	53	2	O37557	human immun
745	5	2.3	53	12	O69872	streptomyce
746	5	2.3	54	12	Q76143	human immun
747	5	2.3	54	12	Q76300	human immun
748	5	2.3	54	12	Q69799	human immun
749	5	2.3	54	12	Q69800	human immun

#					Accession	Description
750	5	2.3	54	12	Q069801	human immun
751	5	2.3	54	12	Q069802	human immun
752	5	2.3	54	12	Q069804	human immun
753	5	2.3	54	12	Q069805	human immun
754	5	2.3	54	12	Q069806	human immun
755	5	2.3	54	12	Q069811	human immun
756	5	2.3	54	12	Q069821	human immun
757	5	2.3	54	12	Q069824	human immun
758	5	2.3	54	12	Q069827	human immun
759	5	2.3	54	12	Q069830	human immun
760	5	2.3	54	12	Q069831	human immun
761	5	2.3	54	12	Q070284	human immun
762	5	2.3	54	12	Q076486	human immun
763	5	2.3	54	12	Q076491	human immun
764	5	2.3	54	12	Q076492	human immun
765	5	2.3	54	12	Q076494	human immun
766	5	2.3	54	12	Q076498	human immun
767	5	2.3	54	12	Q076499	human immun
768	5	2.3	54	12	Q076500	human immun
769	5	2.3	54	12	Q076501	human immun
770	5	2.3	54	12	Q076502	human immun
771	5	2.3	54	12	Q076503	human immun
772	5	2.3	54	12	Q076504	human immun
773	5	2.3	54	12	Q076516	human immun
774	5	2.3	54	12	Q076527	human immun
775	5	2.3	54	12	Q069798	human immun
776	5	2.3	54	12	Q069803	human immun
777	5	2.3	54	12	Q069808	human immun
778	5	2.3	54	12	Q069810	human immun
779	5	2.3	54	12	Q070278	human immun
780	5	2.3	54	12	Q070279	human immun
781	5	2.3	54	12	Q070280	human immun
782	5	2.3	54	12	Q070281	human immun
783	5	2.3	54	12	Q070282	human immun
784	5	2.3	54	12	Q076488	human immun
785	5	2.3	54	12	Q076489	human immun
786	5	2.3	54	12	Q076490	human immun
787	5	2.3	54	12	Q076495	human immun
788	5	2.3	54	12	Q076505	human immun
789	5	2.3	54	12	Q076506	human immun
790	5	2.3	54	12	Q076517	human immun
791	5	2.3	54	12	Q076526	human immun
792	5	2.3	54	12	O37466	human immun
793	5	2.3	54	12	O37467	human immun
794	5	2.3	56	10	Q41836	zea mays (m
795	5	2.3	56	12	O37455	human immun
796	5	2.3	56	12	O37458	human immun
797	5	2.3	56	12	O37462	human immun
798	5	2.3	56	12	O37465	human immun
799	5	2.3	56	12	O37469	human immun
800	5	2.3	56	12	O37471	human immun
801	5	2.3	56	12	O37472	human immun
802	5	2.3	56	12	O37476	human immun
803	5	2.3	56	12	O37478	human immun
804	5	2.3	56	12	O37562	human immun
805	5	2.3	57	12	O37791	human immun
806	5	2.3	57	12	O37804	human immun
807	5	2.3	57	12	O37806	human immun
808	5	2.3	57	12	O37810	human immun
809	5	2.3	57	12	O37811	human immun
810	5	2.3	57	12	O37816	human immun
811	5	2.3	57	12	O37817	human immun
812	5	2.3	58	12	Q85772	human immun
813	5	2.3	58	12	Q85782	human immun
814	5	2.3	59	12	P90112	human immun
815	5	2.3	59	12	O40613	hepatitis c
816	5	2.3	59	12	O40614	hepatitis c
817	5	2.3	59	12	O37563	human immun
818	5	2.3	60	10	O49462	arabidopsis
819	5	2.3	60	12	Q85738	human immun
820	5	2.3	60	12	Q85742	human immun
821	5	2.3	60	12	Q85752	human immun
822	5	2.3	60	12	Q85776	human immun

Row	Accession	Description				
823	Q85753	Q85753 human immun	12	60	2.3	5
824	Q85754	Q85754 human immun	12	60	2.3	5
825	037794	037794 human immun	12	60	2.3	5
826	037795	037795 human immun	12	60	2.3	5
827	037796	037796 human immun	12	60	2.3	5
828	Q9YE88	Q9YE88 aeropyrum p	1	61	2.3	5
829	Q85739	Q85739 human immun	12	61	2.3	5
830	Q85775	Q85775 human immun	12	61	2.3	5
831	Q85777	Q85777 human immun	12	61	2.3	5
832	Q85779	Q85779 human immun	12	61	2.3	5
833	Q85812	Q85812 human immun	12	61	2.3	5
834	Q85813	Q85813 human immun	12	61	2.3	5
835	Q85817	Q85817 human immun	12	61	2.3	5
836	Q85823	Q85823 human immun	12	61	2.3	5
837	Q85773	Q85773 human immun	12	61	2.3	5
838	Q85810	Q85810 human immun	12	61	2.3	5
839	010938	010938 human immun	12	61	2.3	5
840	037807	037807 human immun	12	61	2.3	5
841	037809	037809 human immun	12	61	2.3	5
842	Q14311	Q14311 homo sapien	4	62	2.3	5
843	Q16041	Q16041 homo sapien	4	62	2.3	5
844	Q9XSB3	Q9XSB3 canis famil	6	62	2.3	5
845	Q9XJM2	Q9XJM2 bacteriopha	9	62	2.3	5
846	Q85745	Q85745 human immun	12	62	2.3	5
847	Q85748	Q85748 human immun	12	62	2.3	5
848	Q85784	Q85784 human immun	12	62	2.3	5
849	Q85785	Q85785 human immun	12	62	2.3	5
850	037457	037457 human immun	12	62	2.3	5
851	037559	037559 human immun	12	62	2.3	5
852	Q9Y296	Q9Y296 human immun	12	62	2.3	5
853	Q9Y292	Q9Y292 human immun	12	62	2.3	5
854	Q9Y284	Q9Y284 human immun	12	62	2.3	5
855	031968	031968 bacillus su	7	63	2.3	5
856	Q31236	Q31236 mus musculu	2	63	2.3	5
857	064052	064052 bacteriopha	9	63	2.3	5
858	Q75186	Q75186 human immun	12	63	2.3	5
859	Q85747	Q85747 human immun	12	63	2.3	5
860	Q85786	Q85786 human immun	12	63	2.3	5
861	Q85809	Q85809 human immun	12	63	2.3	5
862	Q85818	Q85818 human immun	12	63	2.3	5
863	Q85822	Q85822 human immun	12	63	2.3	5
864	037454	037454 human immun	12	63	2.3	5
865	037459	037459 human immun	12	63	2.3	5
866	037460	037460 human immun	12	63	2.3	5
867	037461	037461 human immun	12	63	2.3	5
868	037463	037463 human immun	12	63	2.3	5
869	037464	037464 human immun	12	63	2.3	5
870	037468	037468 human immun	12	63	2.3	5
871	037474	037474 human immun	12	63	2.3	5
872	037479	037479 human immun	12	63	2.3	5
873	037482	037482 human immun	12	63	2.3	5
874	037483	037483 human immun	12	63	2.3	5
875	037484	037484 human immun	12	63	2.3	5
876	037485	037485 human immun	12	63	2.3	5
877	037486	037486 human immun	12	63	2.3	5
878	037488	037488 human immun	12	63	2.3	5
879	037489	037489 human immun	12	63	2.3	5
880	037490	037490 human immun	12	63	2.3	5
881	037492	037492 human immun	12	63	2.3	5
882	037493	037493 human immun	12	63	2.3	5
883	037494	037494 human immun	12	63	2.3	5
884	037495	037495 human immun	12	63	2.3	5
885	037496	037496 human immun	12	63	2.3	5
886	037498	037498 human immun	12	63	2.3	5
887	037499	037499 human immun	12	63	2.3	5
888	037500	037500 human immun	12	63	2.3	5
889	037501	037501 human immun	12	63	2.3	5
890	037502	037502 human immun	12	63	2.3	5
891	037503	037503 human immun	12	63	2.3	5
892	037504	037504 human immun	12	63	2.3	5
893	037505	037505 human immun	12	63	2.3	5
894	037506	037506 human immun	12	63	2.3	5
895	037507	037507 human immun	12	63	2.3	5
896	037508	037508 human immun	12	63	2.3	5
897	037509	037509 human immun	12	63	2.3	5
898	037510	037510 human immun	12	63	2.3	5
899	037511	037511 human immun	12	63	2.3	5
900	037512	037512 human immun	12	63	2.3	5
901	037513	037513 human immun	12	63	2.3	5
902	037514	037514 human immun	12	63	2.3	5
903	037515	037515 human immun	12	63	2.3	5
904	037516	037516 human immun	12	63	2.3	5
905	037517	037517 human immun	12	63	2.3	5
906	037518	037518 human immun	12	63	2.3	5
907	037519	037519 human immun	12	63	2.3	5
908	037520	037520 human immun	12	63	2.3	5
909	037521	037521 human immun	12	63	2.3	5
910	037522	037522 human immun	12	63	2.3	5
911	037523	037523 human immun	12	63	2.3	5
912	037524	037524 human immun	12	63	2.3	5
913	037525	037525 human immun	12	63	2.3	5
914	037527	037527 human immun	12	63	2.3	5
915	037530	037530 human immun	12	63	2.3	5
916	037533	037533 human immun	12	63	2.3	5
917	037534	037534 human immun	12	63	2.3	5
918	037535	037535 human immun	12	63	2.3	5
919	037536	037536 human immun	12	63	2.3	5
920	037537	037537 human immun	12	63	2.3	5
921	037538	037538 human immun	12	63	2.3	5
922	037539	037539 human immun	12	63	2.3	5
923	037540	037540 human immun	12	63	2.3	5
924	037541	037541 human immun	12	63	2.3	5
925	037542	037542 human immun	12	63	2.3	5
926	037543	037543 human immun	12	63	2.3	5
927	037544	037544 human immun	12	63	2.3	5
928	037545	037545 human immun	12	63	2.3	5
929	037546	037546 human immun	12	63	2.3	5
930	037547	037547 human immun	12	63	2.3	5
931	037548	037548 human immun	12	63	2.3	5
932	037550	037550 human immun	12	63	2.3	5
933	037551	037551 human immun	12	63	2.3	5
934	037552	037552 human immun	12	63	2.3	5
935	037553	037553 human immun	12	63	2.3	5
936	037555	037555 human immun	12	63	2.3	5
937	037558	037558 human immun	12	63	2.3	5
938	037560	037560 human immun	12	63	2.3	5
939	037565	037565 human immun	12	63	2.3	5
940	037571	037571 human immun	12	63	2.3	5
941	037572	037572 human immun	12	63	2.3	5
942	037574	037574 human immun	12	63	2.3	5
943	037575	037575 human immun	12	63	2.3	5
944	037576	037576 human immun	12	63	2.3	5
945	037578	037578 human immun	12	63	2.3	5
946	037579	037579 human immun	12	63	2.3	5
947	037580	037580 human immun	12	63	2.3	5
948	037582	037582 human immun	12	63	2.3	5
949	037583	037583 human immun	12	63	2.3	5
950	037587	037587 human immun	12	63	2.3	5
951	037589	037589 human immun	12	63	2.3	5
952	037590	037590 human immun	12	63	2.3	5
953	037591	037591 human immun	12	63	2.3	5
954	037592	037592 human immun	12	63	2.3	5
955	037593	037593 human immun	12	63	2.3	5
956	037594	037594 human immun	12	63	2.3	5
957	037595	037595 human immun	12	63	2.3	5
958	037596	037596 human immun	12	63	2.3	5
959	037599	037599 human immun	12	63	2.3	5
960	037617	037617 human immun	12	63	2.3	5
961	037771	037771 human immun	12	63	2.3	5
962	037772	037772 human immun	12	63	2.3	5
963	037773	037773 human immun	12	63	2.3	5
964	037774	037774 human immun	12	63	2.3	5
965	037776	037776 human immun	12	63	2.3	5
966	037777	037777 human immun	12	63	2.3	5
967	037778	037778 human immun	12	63	2.3	5

969	5	2.3	037785	63	12	037785	human immun
970	5	2.3	037788	63	12	037788	human immun
971	5	2.3	037793	63	12	037793	human immun
972	5	2.3	037823	63	12	037823	human immun
973	5	2.3	037841	63	12	037841	human immun
974	5	2.3	037842	63	12	037842	human immun
975	5	2.3	037843	63	12	037843	human immun
976	5	2.3	037844	63	12	037844	human immun
977	5	2.3	037845	63	12	037845	human immun
978	5	2.3	037846	63	12	037846	human immun
979	5	2.3	037847	63	12	037847	human immun
980	5	2.3	037848	63	12	037848	human immun
981	5	2.3	037850	63	12	037850	human immun
982	5	2.3	037851	63	12	037851	human immun
983	5	2.3	037852	63	12	037852	human immun
984	5	2.3	037858	63	12	037858	human immun
985	5	2.3	037859	63	12	037859	human immun
986	5	2.3	037860	63	12	037860	human immun
987	5	2.3	037864	63	12	037864	human immun
988	5	2.3	037865	63	12	037865	human immun
989	5	2.3	037869	63	12	037869	human immun
990	5	2.3	037872	63	12	037872	human immun
991	5	2.3	037873	63	12	037873	human immun
992	5	2.3	037874	63	12	037874	human immun
993	5	2.3	037875	63	12	037875	human immun
994	5	2.3	037876	63	12	037876	human immun
995	5	2.3	037877	63	12	037877	human immun
996	5	2.3	Q85787	63	12	Q85787	human immun
997	5	2.3	Q9ZG67	64	2	Q9ZG67	chlamydia t
998	5	2.3	Q32956	64	8	Q32956	pinus thunb
999	5	2.3	Q35890	64	8	Q35890	simulium ny
1000	5	2.3	Q63005	64	11	Q63005	rattus norv

ALIGNMENTS

```
RESULT 1
P90743
ID   P90743;      PRELIMINARY;      PRT;   341 AA.
AC   P90743;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   C10C6.3 PROTEIN.
GN   C10C6.3
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WHITE S.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z83217; CAB05680.1; -.
SQ   SEQUENCE   341 AA;   37185 MW;   30ED388E CRC32;
```

Query Match   3.6%;   Score 8;   DB 5;   Length 34;
Best Local Similarity   100.0%;   Pred. No. 4;
Matches   8;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  113 VQVEGEPD 120
        ||||||||
Db   45 VQVEGEPD 52
```

```
RESULT 2
O83284
ID   O83284;      PRELIMINARY;      PRT;   448 AA.
AC   O83284;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 51.2 KD PROTEIN.
GN   TP0260.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98332770.
RA   FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA   DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA   SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA   KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA   MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA   HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA   VENTER J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA   DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA   SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA   KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA   MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA   HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA   VENTER J.C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE001207; AAC65253.1; -.
DR   TIGR; TP0260; -.
KW   Hypothetical protein.
SQ   SEQUENCE   448 AA;   51247 MW;   62486D07 CRC32;
```

Query Match   3.6%;   Score 8;   DB 2;   Length 448;
Best Local Similarity   100.0%;   Pred. No. 5.1;
Matches   8;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   55 QFTDEAGR 62
        ||||||||
Db  287 QFTDEAGR 294
```

```
RESULT 3
Q20943
ID   Q20943;      PRELIMINARY;      PRT;   1494 AA.
AC   Q20943;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SIMILAR TO THE HUMAN MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN.
GN   F57C12.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
```

RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD , MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FAVELLO T.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U41554; AAA83299.1; -.
DR   HSSP; P13569; 1NBD.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   ATP-binding; Transport.
SQ   SEQUENCE   1494 AA;   166742 MW;   628DF55C CRC32;

Query Match   3.6%;   Score 8;   DB 5;   Length 1494;
Best Local Similarity   100.0%;   Pred. No. 14;
Matches   8;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   44 ILFNKPYD 51
        |||||||
Db   651 ILFNKPYD 658

RESULT   4
Q94137
ID   Q94137;   PRELIMINARY;   PRT;   1525 AA.
AC   Q94137;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MULTIDRUG RESISTANCE RELATED PROTEIN 2.
GN   MRP-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditida; Rhabditina;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BROEKS A., GERRARD B., ALLIKMETS A., DEAN M., PLASTERK R.H.A.;
RL   EMBO J. 0:0-0(0).
DR   EMBL; U62261; AAB07022.1; -.
DR   HSSP; P13569; 1NBD.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR   PFAM; PF00664; ABC_membrane; 2.
DR   PFAM; PF00005; ABC_tran; 2.
KW   ATP-binding; Transport.
SQ   SEQUENCE   1525 AA;   170216 MW;   DC104A75 CRC32;

Query Match   3.6%;   Score 8;   DB 5;   Length 1525;
Best Local Similarity   100.0%;   Pred. No. 15;
Matches   8;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   44 ILFNKPYD 51
        |||||||
Db   696 ILFNKPYD 703

RESULT   5
Q38044
ID   Q38044   PRELIMINARY;   PRT;   66 AA.
AC   Q38044;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   EXCISIONASE.
OS   Bacteriophage phi-11.
OC   Viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90236919.
RA   YE Z.H., BURANEN S.L., LEE C.Y.;
RT   "Sequence analysis and comparison of int and xis genes from
RT   staphylococcal bacteriophages L54a and phi 11.";
RL   J. Bacteriol. 172:2568-2575(1990).
DR   EMBL; M34832; AAA32197.1; -.
SQ   SEQUENCE   66 AA;   7639 MW;   3BB115E4 CRC32;

Query Match   3.2%;   Score 7;   DB 9;   Length 66;
Best Local Similarity   100.0%;   Pred. No. 10;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   204 YTLDSLA 210
       |||||||
Db   54 YTLDSLA 60

RESULT   6
Q51543
ID   Q51543;   PRELIMINARY;   PRT;   288 AA.
AC   Q51543;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   POBR AND POBA GENES.
GN   POBR.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO1C;
RA   ENTSCH B., SQUIRE L., WICKS R.E.;
RL   Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X76994; CAA54301.1; -.
DR   PFAM; PF00165; HTH_2; 1.
DR   PRINTS; PR00032; HTHARAC.
SQ   SEQUENCE   288 AA;   32176 MW;   DDED4F70 CRC32;

Query Match   3.2%;   Score 7;   DB 2;   Length 288;
Best Local Similarity   100.0%;   Pred. No. 37;
Matches   7;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   113 VQVEGEP 119
       |||||||
Db   63 VQVEGEP 69

RESULT   7
P72806
ID   P72806;   PRELIMINARY;   PRT;   312 AA.
AC   P72806;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   HYPOTHETICAL 35.6 KD PROTEIN.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   TABATA S.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 97061201.
RA   KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA   MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA   HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUSO K., OKUMURA S.,
RA   SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA   TABATA S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90900; BAA16821.1; -.
DR   PFAM; PF00535; Glycos_transf_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   312 AA;  35576 MW;  EB3AE5A0 CRC32;

Query Match             3.2%;  Score 7;  DB 2;  Length 312;
Best Local Similarity   100.0%;  Pred. No. 40;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    71 VQGVYAA 77
         |||||||
Db   298 VQGVYAA 304


RESULT 8
O05938
ID   O05938      PRELIMINARY;      PRT;      352 AA.
AC   O05938;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   LIPASE HELPER PROTEIN.
GN   LPWB.
OS   Pseudomonas wisconsinensis.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LMG P-15151;
RA   HAZBON M.H., DUFEL H., CORNELIS P., JAEGER K.E.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LMG P-15151;
RA   ANDRE C., CHARMOILLE L., CORNELIS P., HAZBON M.H.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U88907; AAB53648.1; -.
SQ   SEQUENCE   352 AA;  39574 MW;  149A3E5C CRC32;

Query Match             3.2%;  Score 7;  DB 2;  Length 352;
Best Local Similarity   100.0%;  Pred. No. 44;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   204 YTLDSLA 210
         |||||||
Db   203 YTLDSLA 209


RESULT 9
Q9Y179
ID   Q9Y179      PRELIMINARY;      PRT;      356 AA.
AC   Q9Y179;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   AMINOPHOSPHOLIPID TRANSLOCASE (FRAGMENT).
GN   APLT-1.
OS   Trichomonas vaginalis.
OC   Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC30325;
RA   SHAH P.H., LUSHBAUGH W.B., FINLEY R.W., LI C.L., MOATE M.E.,
RA   MEADE J.C.;
RT   "Characterization of a family of P-type ATPases in Trichomonas
RT   vaginalis.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF145277; AAD37686.1; -.
FT   NON_TER         1      1
FT   NON_TER       356    356
SQ   SEQUENCE   356 AA;  40036 MW;  BB6B88D9 CRC32;

Query Match             3.2%;  Score 7;  DB 5;  Length 356;
Best Local Similarity   100.0%;  Pred. No. 45;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    85 EGLLVLT 91
         |||||||
Db   127 EGLLVLT 133


RESULT 10
Q9ZMQ3
ID   Q9ZMQ3      PRELIMINARY;      PRT;      406 AA.
AC   Q9ZMQ3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE.
GN   JHP0165.
OS   Helicobacter pylori J99.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J99;
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001455; AAD05746.1; -.
SQ   SEQUENCE   406 AA;  47908 MW;  EC11F71A CRC32;

Query Match             3.2%;  Score 7;  DB 2;  Length 406;
Best Local Similarity   100.0%;  Pred. No. 50;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   210 ANGEWRD 216
         |||||||
Db   356 ANGEWRD 362


RESULT 11
Q00206
ID   Q00206      PRELIMINARY;      PRT;      416 AA.
AC   Q00206;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SUBTILISIN-LIKE PROTEASE (EC 3.4.21.14).
```

```
GN   PEPD.
OS   Aspergillus niger.
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
OC   Eurotiales; Trichocomaceae; anamorphic Trichocomaceae; Aspergillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N400;
RX   MEDLINE; 94156181.
RA   JARAI G., BUXTON F.P.;
RT   "Cloning and characterization of the pepD gene of Aspergillus niger
RT   which codes for a subtilisin-like protease.";
RL   Gene 139:51-57(1994).
DR   EMBL; L19059; AAA32703.1; -.
DR   HSSP; P06873; 1BJR.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PFAM; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
KW   Protease; Hydrolase.
SQ   SEQUENCE  416 AA;  43899 MW;  43279E08 CRC32;

Query Match          3.2%;  Score 7;  DB 3;  Length 416;
Best Local Similarity 100.0%;  Pred. No. 51;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    139 LPAGIER 145
          |||||||
Db     75 LPAGIER 81

RESULT 12
086823
ID   086823         PRELIMINARY;       PRT;   444 AA.
AC   086823;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE 4-AMINOBUTYRATE AMINOTRANSFERASE.
GN   GABT.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   MURPHY L., HARRIS D.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA   KINASHI H., HOPWOOD D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL031225; CAA20213.1; -.
DR   HSSP; P16932; 1DGE.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
DR   PFAM; PF00202; aminotran_3; 1.
KW   Transferase; Aminotransferase.
SQ   SEQUENCE  444 AA;  46479 MW;  24050EB1 CRC32;

Query Match          3.2%;  Score 7;  DB 2;  Length 444;
Best Local Similarity 100.0%;  Pred. No. 54;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     85 EGLLVLT 91
          |||||||
Db    403 EGLLVLT 409

RESULT 13
025572
ID   025572         PRELIMINARY;       PRT;   514 AA.
AC   025572;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 58.9 KD PROTEIN.
GN   HP0914.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000601; AAD07966.1; -.
DR   TIGR; HP0914; -.
KW   Hypothetical protein.
SQ   SEQUENCE  514 AA;  58854 MW;  7317B81E CRC32;

Query Match          3.2%;  Score 7;  DB 2;  Length 514;
Best Local Similarity 100.0%;  Pred. No. 61;
Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    128 RNGVTLN 134
          |||||||
Db    326 RNGVTLN 332

RESULT 14
Q9ZKT5
ID   Q9ZKT5         PRELIMINARY;       PRT;   514 AA.
AC   Q9ZKT5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE OUTER MEMBRANE PROTEIN.
GN   JHP0850.
OS   Helicobacter pylori J99.
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J99;
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
```

RL   Nature 397:176-180(1999).
DR   EMBL; AE001514; AAD06428.1; -.
SQ   SEQUENCE   514 AA; 58892 MW; 198C926D CRC32;

Query Match   3.2%; Score 7; DB 2; Length 514;
Best Local Similarity   100.0%; Pred. No. 61;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   128 RNGVTLN 134
          |||||||
Db   326 RNGVTLN 332

RESULT 15
O14407
ID   O14407   PRELIMINARY;   PRT;   710 AA.
AC   O14407;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   ROPY-2.
GN   ROPY-2.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=74-OR23-1VA;
RA   VIERULA J.P., MAIS J.M.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U23425; AAB68457.1; -.
SQ   SEQUENCE   710 AA; 77817 MW; DBF775EA CRC32;

Query Match   3.2%; Score 7; DB 3; Length 710;
Best Local Similarity   100.0%; Pred. No. 81;
Matches   7; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   124 LAKLRNG 130
          |||||||
Db   347 LAKLRNG 353

Search completed: June 10, 2000, 11:30:40
Job time: 335 sec

```
Percent Similarity: 100.000    Percent Identity: 100.000
              Ratio: 1.000             Gaps: 0

alignment_block:
US-09-252-691-7056 x US-08-123-936-182

Align seg 1/1  to: US-08-123-936-182  from: 1  to: 46

   98 AlaArgLeuThrGlnProGly 104
      |||||||||||||||||||||
    3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-475-228A-182

seq_documentation_block:
; Sequence 182, Application US/08475228A
; Patent No. 5869241
; GENERAL INFORMATION:
;   APPLICANT: Edwards, Cynthia A.
;   APPLICANT: Cantor, Charles R.
;   APPLICANT: Andrews, Beth M.
;   APPLICANT: Turin, Lisa M.
;   APPLICANT: Fry, Kirk E.
;   TITLE OF INVENTION: Sequence-Directed DNA Binding
;   TITLE OF INVENTION: Molecules, Compositions and Methods
;   NUMBER OF SEQUENCES: 664
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genelabs Technologies, Inc.
;     STREET: 505 Penobscot Drive
;     CITY: Redwood City
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94063
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/475,228A
;     FILING DATE: 06-JUN-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/123,936
;     FILING DATE: 17-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/996,783
;     FILING DATE: 23-DEC-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/723,618
;     FILING DATE: 27-JUN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/081,070
;     FILING DATE: 22-JUN-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Stratford, Carol A.
;     REGISTRATION NUMBER: 34,444
;     REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 324-0880
;     TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 46 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;     INDIVIDUAL ISOLATE: Human hsp70B gene
US-08-475-228A-182
```

```
alignment_scores:
          Quality: 7.00          Length: 7
            Ratio: 1.000           Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-475-228A-182

Align seg 1/1  to: US-08-475-228A-182  from: 1  to: 46

   98 AlaArgLeuThrGlnProGly 104
      |||||||||||||||||||||
    3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-482-080A-182

seq_documentation_block:
; Sequence 182, Application US/08482080A
; Patent No. 6010849
; GENERAL INFORMATION:
;   APPLICANT: Edwards, Cynthia A.
;   APPLICANT: Cantor, Charles R.
;   APPLICANT: Andrews, Beth M.
;   APPLICANT: Turin, Lisa M.
;   APPLICANT: Fry, Kirk E.
;   TITLE OF INVENTION: Sequence-Directed DNA Binding
;   TITLE OF INVENTION: Molecules, Compositions and Methods
;   NUMBER OF SEQUENCES: 664
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genelabs Technologies, Inc.
;     STREET: 505 Penobscot Drive
;     CITY: Redwood City
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94063
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/482,080A
;     FILING DATE: 07-JUN-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/171,389
;     FILING DATE: 20-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/123,936
;     FILING DATE: 17-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/996,783
;     FILING DATE: 23-DEC-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/723,618
;     FILING DATE: 27-JUN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/081,070
;     FILING DATE: 22-JUN-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Brady, John F.
;     REGISTRATION NUMBER: 39,118
;     REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (650) 324-0880
;     TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 46 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
```

```
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:    Human hsp70B gene
US-08-482-080A-182

alignment_scores:
           Quality:    7.00              Length:    7
           Ratio:     1.000              Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-482-080A-182   ..

Align seg 1/1   to: US-08-482-080A-182   from: 1   to: 46

    98 AlaArgLeuThrGlnProGly 104
       |||||||||||||||||||||
     3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-12388-182

seq_documentation_block:
; Sequence 182, Application PC/TUS9312388
; GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:    Sequence-Directed DNA Binding
;     TITLE OF INVENTION:    Molecules, Compositions and Methods
;     NUMBER OF SEQUENCES:   641
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:   Genelabs Technologies, Inc.
;       STREET:      505 Penobscot Drive
;       CITY:        Redwood City
;       STATE:       CA
;       COUNTRY:     USA
;       ZIP:         94063
; COMPUTER READABLE FORM:
;     MEDIUM TYPE:       Floppy disk
;     COMPUTER:          IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:          Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/123,936
;     FILING DATE:         17-SEP-1993
; PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/996,783
;     FILING DATE:         23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
;     NAME:  Fabian, Gary R.
;     REGISTRATION NUMBER:  33,875
;     REFERENCE/DOCKET NUMBER:  4600-0175.41/G19PCT2
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 324-0880
;     TELEFAX:    (415) 324-0960
; INFORMATION FOR SEQ ID NO: 182:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  46 base pairs
;       TYPE:    nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  Human hsp70B gene
PCT-US93-12388-182

alignment_scores:
           Quality:    7.00              Length:    7
           Ratio:     1.000              Gaps:      0
```

```
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x PCT-US93-12388-182   ..

Align seg 1/1   to: PCT-US93-12388-182   from: 1   to: 46

    98 AlaArgLeuThrGlnProGly 104
       |||||||||||||||||||||
     3 GCCCGGCTGACTCAGCCCGGG 23

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-468-819-73

seq_documentation_block:
; Sequence 73, Application US/08468819
; Patent No. 5871723
; GENERAL INFORMATION:
;     APPLICANT:  Strieter, Robert M.
;     APPLICANT:  Polverini, Peter J.
;     APPLICANT:  Kunkel, Steven L.
;     TITLE OF INVENTION:  CXC Chemokines as Regulators of
;     TITLE OF INVENTION:  Angiogenesis
;     NUMBER OF SEQUENCES:  93
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:     P.O. Box 4433
;       CITY:       Houston
;       STATE:      TX
;       COUNTRY:    US
;       ZIP:        77210
; COMPUTER READABLE FORM:
;     MEDIUM TYPE:       Floppy disk
;     COMPUTER:          IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:          Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/468,819
;     FILING DATE:         Concurrently herewith
;     CLASSIFICATION:      424
; ATTORNEY INFORMATION:
;     NAME:  Highlander, Steven L.
;     REGISTRATION NUMBER:  37,642
;     REFERENCE/DOCKET NUMBER:  UMIC:003/HYL
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  512/418-3000
;     TELEFAX:    512/474-7477
;     TELEX:      N/A
; INFORMATION FOR SEQ ID NO: 73:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  654 base pairs
;       TYPE:    nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
US-08-468-819-73

alignment_scores:
           Quality:    7.00              Length:    7
           Ratio:     1.000              Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-468-819-73/rev   ..

Align seg 1/1   to reverse of: US-08-468-819-73   from: 1   to: 654

   157 ProProIleArgGluArgLys 163
       |||||||||||||||||||||
   196 CCACCTATCAGGGAACGGGAAG 176

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-468-819-75
```

seq_documentation_block:
; Sequence 75, Application US/08468819
; Patent No. 5871723
; GENERAL INFORMATION:
;   APPLICANT: Strieter, Robert M.
;   APPLICANT: Polverini, Peter J.
;   APPLICANT: Kunkel, Steven L.
;   TITLE OF INVENTION: CXC Chemokines as Regulators of
;   TITLE OF INVENTION: Angiogenesis
;   NUMBER OF SEQUENCES: 93
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Arnold, White & Durkee
;   STREET: P.O. Box 4433
;   CITY: Houston
;   STATE: TX
;   COUNTRY: US
;   ZIP: 77210
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/468,819
;   FILING DATE: Concurrently herewith
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Highlander, Steven L.
;   REGISTRATION NUMBER: 37,642
;   REFERENCE/DOCKET NUMBER: UMIC:003/HYL
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 512/418-3000
;   TELEFAX: 512/474-7477
;   TELEX: N/A
; INFORMATION FOR SEQ ID NO: 75:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 654 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   DESCRIPTION: /desc = "DNA"
US-08-468-819-75

alignment_scores:
       Quality: 7.00   Length: 7
       Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-468-819-75/rev ..

Align seq 1/1 to reverse of: US-08-468-819-75   from: 1 to: 654

    157 ProProIleArgGluArgLys 163
        ||||||||||||||||||||
    196 .CCACCTATCAGGGAACGGAAG 176

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-997-080-40

seq_documentation_block:
; Sequence 40, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
;   APPLICANT: WATSON, JAMES D.
;   APPLICANT: TAN, PAUL L.J.
;   TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
;   NUMBER OF SEQUENCES: 194
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Law Offices of Ann W. Speckman
;   STREET: 2601 Elliott Avenue, Suite 4185

;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98121
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/997,080
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sleath, Janet
;   REGISTRATION NUMBER: 37,007
;   REFERENCE/DOCKET NUMBER: 11000.1007
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-269-0565
;   TELEFAX: 206-269-0563
;   TELEX:
; INFORMATION FOR SEQ ID NO: 40:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1211 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
US-08-997-080-40

alignment_scores:
       Quality: 7.00   Length: 7
       Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-997-080-40/rev ..

Align seq 1/1 to reverse of: US-08-997-080-40   from: 1 to: 1211

    76 AlaAlaGlyArgLeuAspArg 82
       |||||||||||||||||||||
   507 GCTGCCGGTCGGCTTGACCGC 487

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-997-362-40

seq_documentation_block:
; Sequence 40, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
;   APPLICANT: Tan, Paul
;   APPLICANT: Hiyama, Jun
;   APPLICANT: Visser, Elizabeth
;   APPLICANT: Skinner, Margot
;   APPLICANT: Scott, Linda
;   APPLICANT: Prestidge, Ross
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
;   TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;   NUMBER OF SEQUENCES: 194
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Law Offices of Ann W. Speckman
;   STREET: 2601 Elliott Avenue, Suite 4185
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98121
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible

```
        OPERATING SYSTEM: DOS
        SOFTWARE: FastSEQ for Windows Version 2.0
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/997,362
        FILING DATE:
        CLASSIFICATION:
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
        FILING DATE: June 12, 1997
        APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
        FILING DATE: August 29, 1996
    ATTORNEY/AGENT INFORMATION:
        NAME: Sleath, Janet
        REGISTRATION NUMBER: 37,007
        REFERENCE/DOCKET NUMBER: 11000.1002c2
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 206-269-0565
        TELEFAX: 206-269-0563
        TELEX:
INFORMATION FOR SEQ ID NO: 40:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1211 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
        MOLECULE TYPE: Genomic DNA
US-08-997-362-40

alignment_scores:
        Quality:      7.00          Length:      7
        Ratio:        1.000         Gaps:        0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-997-362-40/rev  ..

Align seg 1/1  to reverse of: US-08-997-362-40  from: 1  to: 1211

    76 AlaAlaGlyArgLeuAspArg 82
       |||||||||||||||||||||||
   507 GCTGCCGGTCGGCTTGACCGC 487

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-873-970-40

seq_documentation_block:
; Sequence 40, Application US/08873970
; Patent No. 6001361
    GENERAL INFORMATION:
        APPLICANT: Tan, Paul
        APPLICANT: Hiyama, Jun
        APPLICANT: Visser, Elizabeth
        APPLICANT: Skinner, Margot
        APPLICANT: Scott, Linda
        APPLICANT: Prestidge, Ross
        TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
        TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
        NUMBER OF SEQUENCES: 106
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Law Offices of Ann W. Speckman
        STREET: 2601 Elliott Avenue, Suite 4185
        CITY: Seattle
        STATE: WA
        COUNTRY: USA
        ZIP: 98121
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Diskette
        COMPUTER: IBM Compatible
        OPERATING SYSTEM: DOS
        SOFTWARE: FastSEQ for Windows Version 2.0
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/873,970
        FILING DATE:
```

```
        CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 08/705,347
        FILING DATE: 29-AUG-1996
    ATTORNEY/AGENT INFORMATION:
        NAME: Sleath, Janet
        REGISTRATION NUMBER: 37,007
        REFERENCE/DOCKET NUMBER: 11000.1002C1
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 206-269-0565
        TELEFAX: 206-269-0563
        TELEX:
INFORMATION FOR SEQ ID NO: 40:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1211 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
        MOLECULE TYPE: Genomic DNA
US-08-873-970-40

alignment_scores:
        Quality:      7.00          Length:      7
        Ratio:        1.000         Gaps:        0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-873-970-40/rev  ..

Align seg 1/1  to reverse of: US-08-873-970-40  from: 1  to: 1211

    76 AlaAlaGlyArgLeuAspArg 82
       |||||||||||||||||||||||
   507 GCTGCCGGTCGGCTTGACCGC 487

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-468-709B-5

seq_documentation_block:
; Sequence 5, Application US/08468709B
; Patent No. 5654137
    GENERAL INFORMATION:
        APPLICANT: Astrom, Anders
        APPLICANT: Voorhees, John
        APPLICANT: Pattersson, Ulrika
        APPLICANT: Tavakkol, Amir
        TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
        NUMBER OF SEQUENCES: 18
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
        STREET: PO Box 828
        CITY: Bloomfield Hills
        STATE: Michigan
        COUNTRY: United States of America
        ZIP: 48303
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/468,709B
        FILING DATE: 06/06/95
        CLASSIFICATION: 435
    ATTORNEY/AGENT INFORMATION:
        NAME: Smith, Deann F.
        REGISTRATION NUMBER: 36,683
        REFERENCE/DOCKET NUMBER: 2115-00676DVF
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (810) 641-1600
        TELEFAX: (810) 641-0270
        TELEX: 287637
INFORMATION FOR SEQ ID NO: 5:
```

```
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 1322 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;      ORGANISM: Homo sapiens
;      TISSUE TYPE: Placenta
;   IMMEDIATE SOURCE:
;      LIBRARY: human placenta genomic library
;      CLONE: lambda 2.1
;   FEATURE:
;      NAME/KEY: TATA_signal
;      LOCATION: 1008..1013
;   FEATURE:
;      NAME/KEY: exon
;      LOCATION: 1039..1245
;   FEATURE:
;      NAME/KEY: intron
;      LOCATION: 1246..1322
;   PUBLICATION INFORMATION:
;      AUTHORS: Astrom, Anders
;      AUTHORS: Pettersson, Ulrika
;      AUTHORS: Voorhees, John J
;      TITLE: Structure of the human cellular retinoic
;      TITLE: acid-binding protein II (CRABP-II) gene: Early
;      TITLE: transcriptional regulation by retinoic acid
;      JOURNAL: J. Biol. Chem.
;      DATE: 1992
;      RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 1322
US-08-468-709B-5

alignment_scores:
;      Quality:      7.00          Length:       7
;      Ratio:        1.000          Gaps:         0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-468-709B-5/rev    ..

Align seg 1/1  to reverse of: US-08-468-709B-5   from: 1   to: 1322

         27 SerSerArgGlnAlaThrArg 33
            |||||||||||||||||||||
        399 AGTAGTCGCCAGGCGACTCGC 379

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-241-664B-5
seq_documentation_block:
; Sequence 5, Application US/08241664B
Patent No. 5871909
GENERAL INFORMATION:
;      APPLICANT: Voorhees, John J.
;      APPLICANT: Astrom, Anders
;      APPLICANT: Pettersson, Ulrika
;      APPLICANT: Tavakkol, Amir
;   TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
;      STREET: PO Box 828
;      CITY: Bloomfield Hills
;      STATE: Michigan
;      COUNTRY: United States of America
;      ZIP: 48303
;   COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/241,664B
;      FILING DATE: May 11, 1994
;      CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;      NAME: Smith, Deann F.
;      REGISTRATION NUMBER: 36,683
;      REFERENCE/DOCKET NUMBER: 2115-0067COD
;   TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (810) 641-1600
;      TELEFAX: (810) 641-0270
;      TELEX: 287637
;   INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;      LENGTH: 1322 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;      ORGANISM: Homo sapiens
;      TISSUE TYPE: Placenta
;   IMMEDIATE SOURCE:
;      LIBRARY: human placenta genomic library
;      CLONE: lambda 2.1
;   FEATURE:
;      NAME/KEY: TATA_signal
;      LOCATION: 1008..1013
;   FEATURE:
;      NAME/KEY: exon
;      LOCATION: 1039..1245
;   FEATURE:
;      NAME/KEY: intron
;      LOCATION: 1246..1322
;   PUBLICATION INFORMATION:
;      AUTHORS: Astrom, Anders
;      AUTHORS: Pettersson, Ulrika
;      AUTHORS: Voorhees, John J
;      TITLE: Structure of the human cellular retinoic
;      TITLE: acid-binding protein II (CRABP-II) gene: Early
;      TITLE: transcriptional regulation by retinoic acid
;      JOURNAL: J. Biol. Chem.
;      DATE: 1992
;      RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 1322
US-08-241-664B-5

alignment_scores:
;      Quality:      7.00          Length:       7
;      Ratio:        1.000          Gaps:         0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-241-664B-5/rev    ..

Align seg 1/1  to reverse of: US-08-241-664B-5   from: 1   to: 1322

         27 SerSerArgGlnAlaThrArg 33
            |||||||||||||||||||||
        399 AGTAGTCGCCAGGCGACTCGC 379

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US93-03936-5
seq_documentation_block:
; Sequence 5, Application PC/TUS9303936
GENERAL INFORMATION:
;      APPLICANT: Voorhees, John J.
;      APPLICANT: Astrom, Anders
;      APPLICANT: Pettersson, Ulrika
;      APPLICANT: Tavakkol, Amir
```

TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce
STREET: PO Box 828
CITY: Bloomfield Hills
STATE: Michigan
COUNTRY: United States of America
ZIP: 48013
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03936
FILING DATE: 19930427
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lewak, Anna M.
REGISTRATION NUMBER: 33,006
REFERENCE/DOCKET NUMBER: 21150676POB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (313) 641-1600
TELEFAX: (313) 641-0270
TELEX: 287637
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1322 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: placenta
IMMEDIATE SOURCE:
LIBRARY: human placenta genomic library
CLONE: lambda 2.1
FEATURE:
NAME/KEY: TATA_signal
LOCATION: 1008..1013
FEATURE:
NAME/KEY: exon
LOCATION: 1039..1245
FEATURE:
NAME/KEY: intron
LOCATION: 1246..1322
PUBLICATION INFORMATION:
AUTHORS: Astrom, Anders
AUTHORS: Pettersson, Ulrika
AUTHORS: Voorhees, John J
TITLE: structure of the human cellular retinoic
TITLE: acid-binding protein II (CRABP-II) gene: Early
TITLE: transcriptional regulation by retinoic acid
JOURNAL: J. Biol. Chem.
DATE: 1992
RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 1322
PCT-US93-03936-5

alignment_scores:
Quality: 7.00   Length: 7
Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x PCT-US93-03936-5/rev ..

Align seg 1/1 to reverse of: PCT-US93-03936-5 from: 1 to: 1322

---

27 SerSerArgGlnAlaThrArg 33
    |||||||||||||||||||||
399 AGTAGTCGCCAGGCGACTCGC 379

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-389-812-3

seq_documentation_block:
; Sequence 3, Application US/08389812
; Patent No. 6015665
GENERAL INFORMATION:
APPLICANT: DIXIT, VISHVA M.
TITLE OF INVENTION: METHOD AND COMPOSITION FOR REGULATING
TITLE OF INVENTION: APOPTOSIS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,812
FILING DATE: 13-FEB-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: KONSKI, ANTOINETE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20344-21036.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1468 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 295..1317
US-08-389-812-3

alignment_scores:
Quality: 7.00   Length: 7
Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-389-812-3 ..

Align seg 1/1 to: US-08-389-812-3 from: 1 to: 1468

127 LeuArgAsnGlyValThrLeu 133
    |||||||||||||||||||||
950 TTAAGAAATGGTGTGACTCTA 970

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-583-318-4

seq_documentation_block:
; Sequence 4, Application US/08583318
; Patent No. 5693483
GENERAL INFORMATION:
APPLICANT: Staunton, Donald
APPLICANT: Harris, Edith
TITLE OF INVENTION: Cytoplasmic Modulators of Integrin

```
TITLE OF INVENTION:  Binding
NUMBER OF SEQUENCES:  14
CORRESPONDENCE ADDRESS:
  ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
  STREET:  233 South Wacker Drive, 6300 Sears Tower
  CITY:  Chicago
  STATE:  Illinois
  COUNTRY:  USA
  ZIP:  60606
COMPUTER READABLE FORM:
  MEDIUM TYPE:  Floppy disk
  COMPUTER:  IBM PC compatible
  OPERATING SYSTEM:  PC-DOS/MS-DOS
  SOFTWARE:  PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER:  US/08/583,318
  FILING DATE:
  CLASSIFICATION:  435
ATTORNEY/AGENT INFORMATION:
  NAME:  Williams Jr., Joseph A.
  REGISTRATION NUMBER:  38,659
  REFERENCE/DOCKET NUMBER:  27866/33050
TELECOMMUNICATION INFORMATION:
  TELEPHONE:  312-474-6300
  TELEFAX:  312-474-0448
INFORMATION FOR SEQ ID NO:  4:
SEQUENCE CHARACTERISTICS:
  LENGTH:  1631 base pairs
  TYPE:  nucleic acid
  STRANDEDNESS:  single
  TOPOLOGY:  linear
MOLECULE TYPE:  CDNA
FEATURE:
  NAME/KEY:  CDS
  LOCATION:  1..1437
US-08-583-318-4
```

```
alignment_scores:
          Quality:     7.00                    Length:     7
          Ratio:     1.000                      Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x US-08-583-318-4/rev  ..

Align seg 1/1   to reverse of: US-08-583-318-4   from: 1   to: 1631

   124 LeuAlaLysLeuArgAsnGly 130
       |||||||||||||||||||||
   488 TTGGCAAAACTGCCGAATGGC 468
```

OM of: US-09-252-691-7056 to: EST:*   out_format : pfs

Date: Jun 10, 2000 2:23 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09252691/runat_05062000_101737_1252/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -GAPOP=4.500
-GAPEXT=-0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=-0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=-7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cd1 -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=-1 -ALIGN=-15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09252691 -NCPU=6 -ICPU=3 -NO_XLPXY -WAIT
-THREADS=1

Search information block:
Query: US-09-252-691-7056
Query length: 222
Database: EST:*
Database sequences: 4857316
Database length: 202611650
Search time (sec): 604.990000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est26:AI373988	+	52.00	990.63	2.2e-46	283	! AI373988 SWOvAFCAP32B02SK Oncho
em_est20:AW017830	+	25.00	467.99	2.8e-17	295	! Aw017830 614064A09.x1 614 - roc
gb_gss11:AQ212974	-	10.00	173.41	0.7289	546	! AQ212974 HS_3214_B2_E03_MR CIT
gb_est26:AA237292	-	8.00	143.47	33.90	160	! AA237292 mw96d12.r1 Soares mous
gb_est19:AA755409	-	8.00	140.82	47.63	232	! AA755409 vr53e06.r1 Knowles Sol
gb_est39:AV206648	+	8.00	140.35	50.62	248	! AV206648 AV206648 RIKEN full-le
gb_est12:AA286664	+	8.00	140.32	50.81	249	! AA286664 vb79h04.r1 Soares mous
gb_est18:AA688759	+	8.00	140.12	52.11	256	! AA688759 zm10c07.r1 Knowles Sol
gb_est1:X73795	-	8.00	139.71	54.90	271	! X73795 CHESTM092 Goat mammary g
gb_est27:AI424418	+	8.00	139.21	58.59	291	! AI424418 te90f02.x1 NCI_CGAP_Pr
gb_est42:AV309318	+	8.00	138.89	60.98	304	! AV309318 AV309318 RIKEN full-le
gb_est41:AI047546	-	8.00	138.14	69.37	338	! AI047546 uh80b05.r1 Soares mous
gb_est26:AI329557	+	8.00	137.89	69.37	350	! AI329557 b6c08ne.f1 Neurospora
gb_est38:AW047543	-	8.00	137.75	70.64	357	! AW047543 UI-M-BH1-ama-f-02-0-UI
gb_est11:T52062	+	8.00	137.42	73.71	374	! T52062 yb29d04.r1 Stratagene fe
gb_est9:AA082959	-	8.00	137.05	77.31	394	! AA082959 zn07b11.r1 Stratagene
gb_est35:AA069085	+	8.00	136.97	78.02	398	! AA069085 zm10c07.r1 Stratagene
gb_est39:AW123360	+	8.00	136.88	78.92	403	! D27082 CELK008C6F Yuji Kohara u
gb_est9:AA098014	+	8.00	136.85	79.28	405	! AA098014 mn85e11.r1 Stratagene
gb_est35:AL040306	-	8.00	136.62	81.60	418	! AL040306 DKFZp434M1013_r1 434 -
gb_est39:AW123360	-	8.00	136.57	82.14	421	! AW123360 UI-M-BH2.1-apf-e-10-0-
gb_est27:AI398322	-	8.00	136.51	82.85	425	! AI398322 NCSM2H9T3 Subtracted M
gb_est25:AI273549	+	8.00	136.26	85.52	440	! AI273549 q155d03.x1 Soares_NhHM
gb_est6:W09815	+	8.00	136.13	86.94	448	! W09815 ma49e07.r1 Soares mouse
gb_gss4:AQ687042	-	8.00	136.07	87.65	452	! AQ687042 nbxb0073N01r CUGI Rice
gb_est5:W09817		8.00	136.03	88.01	454	! W09817 ma49f01.r1 Soares mouse
gb_est7:W30249	+	8.00	135.91	89.42	462	! W30249 mc28b01.r1 Soares mouse
gb_est23:AI140187	-	8.00	135.79	90.84	470	! AI140187 qe20b09.x1 Soares_feta
gb_est19:AA793606	-	8.00	135.56	93.49	485	! AA793606 vp79e08.r1 Stratagene
gb_est11:AA250690	-	8.00	135.49	94.37	490	! AA250690 mx79h02.r1 Stratagene mous
gb_est9:AA072069	-	8.00	135.42	95.25	495	! AA072069 mm69g06.r1 Stratagene
gb_est11:AA270914	+	8.00	135.42	95.25	495	! AA270914 va70c06.r1 Soares mous
gb_est11:AA060088	-	8.00	135.29	96.83	504	! AA060088 mj71b09.r1 Soares mous
gb_est18:AA674188	-	8.00	135.16	98.41	513	! AA674188 vp97c02.r1 Stratagene
gb_est25:AI329555	+	8.00	135.14	98.76	515	! AI329555 b6b1ne.f1 Neurospora
gb_gss5:AQ789052	+	8.00	135.14	98.76	515	! AQ789052 HS_3208_B2_D04_MR CIT
gb_est37:AW012406	-	8.00	135.08	99.46	519	! AW012406 uo04e08.y1 Sugano mous
gb_gss13:AQ367068	-	8.00	135.04	99.99	522	! AQ367068 HS_5032_B1_F05_T7 RPCI
gb_est1:AA086716	-	8.00	134.89	101.91	533	! AA086716 mn93h11.r1 Stratagene

Documentation	Sequence	Len	EScore	ZScore	Orig	Strd
! AA516785 vh88d10.r1 Knowles	gb_est15:AA516785	533	101.91	134.89	8.00	-
! AA268935 va99g10.r1 Soares m	gb_est11:AA268935	534	102.09	134.88	8.00	-
! AA266576 mz53e10.r1 Barstead	gb_est11:AA266576	535	102.26	134.86	8.00	-
! AW211779 uo85e01.y1 NCI_CGAP	gb_est11:AW211779	535	102.26	134.86	8.00	-
! AI786062 uj57h02.y1 Sugano m	gb_est34:AI786062	545	104.01	134.73	8.00	+
! AA475027 vh03f09.r1 Soares_m	gb_est15:AA475027	552	105.23	134.64	8.00	-
! AI182222 uc68c04.r1 Soares_m	gb_est14:AI182222	553	105.41	134.63	8.00	-
! AA445266 vf62h11.r1 Barstead	gb_est14:AA445266	553	105.41	134.63	8.00	-
! AI954969 wx92b02.x1 NCI_CGAP	gb_est37:AI954969	567	107.84	134.45	8.00	-
! AI853843 UI-M-BH0-ajr-d-02-0	gb_est1:AI853843	570	108.37	134.41	8.00	+
! AQ972312 RPCI-23-312L19.TV R	gb_gss7:AQ972312	570	108.37	134.41	8.00	-
! AI035497 ub48b04.r1 Soares_m	gb_est22:AI035497	574	109.06	134.36	8.00	-
! AQ273411 nbxb0003E21r CUGI R	gb_gss11:AQ273411	582	110.45	134.25	8.00	-
! AQ025734 l(2)k0003 Drosophi	gb_gss9:AQ025734	583	110.62	134.25	8.00	-
! AQ948077 Sheared DNA-50N7.TF	gb_gss7:AQ948077	592	112.19	134.14	8.00	+
! AQ363138 nbxb005A20f CUGI R	gb_gss7:AQ363138	596	112.88	134.09	8.00	-
! AI529215 u16lc06.y1 Sugano m	gb_est28:AI529215	603	114.09	134.01	8.00	-
! AA869364 vq07b11.r1 Barstead	gb_est20:AA869364	613	115.82	133.89	8.00	-
! AI303127 u160b05.y1 Sugano m	gb_est25:AI303127	641	120.65	133.37	8.00	-
! AI226044 uj08f03.y1 Sugano m	gb_est24:AI226044	660	123.92	133.37	8.00	-
! AG010516 Homo sapiens genomi	gb_gss1:AG010516	722	134.52	132.73	8.00	-
! AG010526 Homo sapiens genomi	gb_gss1:AG010526	731	136.05	132.64	8.00	-
! AI573796 uj65h03.y1 Sugano m	gb_est29:AI573796	783	144.88	132.15	8.00	-
! AI119391 uf03f07.y1 Sugano m	gb_est23:AI119391	828	152.48	131.75	8.00	+
! B11311 F22P19-Sp6 IGF Arabid	gb_gss3:B11311	897	164.06	131.36	8.00	-
! AI040945 ov53b1l.x1 Soares_te	gb_est22:AI040945	58	160.24	131.36	7.00	-
! AW177774 RCO-HT0023-180899-0	gb_est44:AW177774	107	280.56	126.99	7.00	-
! AI992098 ws21g10.x1 NCI_CGAP	gb_est37:AI992098	110	287.74	126.80	7.00	-
! T10501 hbc553 Human pancreat	gb_est1:T10501	122	316.32	126.06	7.00	-
! AI612921 tx16g01.x1 NCI_CGAP	gb_est29:AI612921	128	330.52	125.72	7.00	-
! AI096038 SWOvL3CAN20E03 Onch	gb_est23:AI096038	134	344.66	125.29	7.00	-
! AV087819 Mus muscul	gb_est33:AV087819	136	349.37	125.28	7.00	-
! AW031720 EST275174 tomato ca	gb_est38:AW031720	138	354.06	125.18	7.00	-
! AW176501 RC4-CT0036-180899-0	gb_est44:AW176501	155	393.75	124.23	7.00	+
! AA846992 oe08g09.s1 NCI_CGAP	gb_est20:AA846992	169	426.16	123.74	7.00	-
! U38048 OSU38048 FDRSC Oryza	gb_gss5:U38048	171	430.77	123.65	7.00	-
! AQ212974 HS_3214_B2_E03_MR CIT	gb_gss13:AQ212974	189	472.06	122.94	7.00	-
! AA727187 vu91c11.r1 Stratage	gb_est18:AA727187	191	476.63	122.86	7.00	-
! AA218764 SWOvAFCAP12B07SK On	gb_est18:AA218764	193	481.19	122.79	7.00	+
! AW253159 UI-R-BJO-adv-b-09-0	gb_est45:AW253159	193	481.19	122.79	7.00	+
! AA620260 SWOvL3CA655SK Oncho	gb_est17:AA620260	205	508.48	122.36	7.00	+
! AA928279 RPCI-23-262D8.TJ RP	gb_est11:AA928279	206	510.75	122.32	7.00	-
! AQ256663 ty58g05.x1 NCI_CGAP	gb_gss7:AQ256663	208	515.29	122.32	7.00	-
! C29055 C29055 Rice callus cD	gb_est1:C29055	211	522.08	122.15	7.00	-
! AQ467486 HS_5219_A2_C05_SP6E	gb_gss14:AQ467486	212	524.34	122.12	7.00	-
! AV150321 SWOvL3CAN20E03 Onch	gb_est34:AV150321	217	535.64	121.85	7.00	-
! W33989 mb57e04.r1 Soares mou	gb_est7:W33989	220	542.41	121.79	7.00	-
! AW176774 RC0-CT0094-280899-0	gb_est44:AW176774	222	546.92	121.69	7.00	-
! AW346666 AV346666 RIKEN full	gb_est43:AW346666	223	549.17	121.54	7.00	-
! R73420 yj192h10.r1 Soares bre	gb_est3:R73420	225	553.67	121.48	7.00	-
! AV377646 yj192h10.r1 Soares m	gb_est44:AV377646	230	564.92	121.45	7.00	-
! AV287139 4784 MARC 1PIG Sus	gb_est44:AV287139	232	569.41	121.35	7.00	-
! AI362027 EST71371 T-cell lym	gb_est13:AI362027	235	576.14	121.32	7.00	-
! AI253745 aq28f07.x1 Barstead	gb_est25:AI253745	236	578.38	121.23	7.00	-
! AA299011 EST11467 Uterus Hom	gb_est12:AA299011	237	580.62	121.23	7.00	-
! AA201681 q144a05.x1 NCI_CGAP	gb_est24:AA201681	239	585.10	121.23	7.00	-
! Z42139 HSCOAA101 normalized	gb_est1:Z42139	240	587.34	121.03	7.00	-
! AV503961 vm40d06.x1 Knowles	gb_est8:AV503961	241	589.58	121.00	7.00	-
! AV362827 RIKEN full	gb_est43:AV362827	247	602.99	121.00	7.00	-
! AV341813 AV341813 RIKEN full	gb_est43:AV341813	248	605.22	121.00	7.00	-
! AA015004 mh26a03.r1 Soares m	gb_est8:AA015004	248	605.22	120.91	7.00	-
! AV173917 AV173917 Mus muscul	gb_est34:AV173917	251	611.91	120.89	7.00	-
! C23540 C23540 Japanese floun	gb_est16:C23540	252	614.14	120.86	7.00	-
! AQ107318 HS_3102_A2_G10_MR C	gb_gss10:AQ107318	253	616.37	120.83	7.00	-
! AW437009 77688 MARC 2PIG.Sus	gb_est47:AW437009	254	618.60	120.75	7.00	-
! AA980755 GM210006B20F3R Gm-r	gb_est21:AA980755	257	625.28	120.75	7.00	+
! AW431479 72057 MARC 2PIG Sus	gb_est47:AW431479	260	631.95	120.66	7.00	-
! AV268759 AV268759 RIKEN full	gb_est41:AV268759	263	638.62	120.58	7.00	-
! Z40954 HSC2KB092 normalized	gb_est1:Z40954	264	640.84	120.55	7.00	-

Left block (ranks 266–325)

Strand	Oligo ID		Pos	Tm	Rank	Accession / Clone
+	gb_est1:T34114	7.00	645.27	120.50	266	T34114 EST62808 Human White bld
+	gb_est31:AV029411	7.00	654.14	120.39	270	AV029411 AV029411 Mus musculus
+	gb_est43:AV344204	7.00	654.14	120.34	270	AV344204 AV344204 RIKEN full-le
+	gb_est41:AV270638	7.00	658.57	120.26	272	AV270638 AV270638 RIKEN full-le
-	gb_est18:AA684340	7.00	665.21	120.26	275	AA684340 vm69c05.s1 Knowles Sol
-	gb_est10:AQ097520	7.00	665.21	120.24	276	AQ097520 HS_3038_B1_F09_MR CIT
-	gb_est3:R73216	7.00	667.43	120.19	276	R73216 yj91h08.r1 Soares breast
-	gb_est2:AI129183	7.00	671.85	120.19	278	AI129183 ok31h09.s1 Soares_NSF
+	gb_est31:AV038986	7.00	671.85	120.19	278	AV038986 AV038986 Mus musculus
+	gb_est40:AV222987	7.00	671.85	120.16	278	AV222987 AV222987 RIKEN full-le
-	gb_est44:AW228456	7.00	674.06	120.14	279	AW228456 up23a08.y1 NCI_CGAP_Ma
+	gb_est27:AI426099	7.00	676.27	120.14	280	AI426099 mh38f04.x1 Soares mous
-	gb_est18:AA721235	7.00	676.22	120.11	280	AA721235 n272b10.x1 NCI_CGAP_GC
-	gb_est30:AI666302	7.00	678.48	120.08	281	AI666302 mt65f09.x1 Soares mous
+	gb_est32:AV060268	7.00	680.68	120.08	282	AV060268 AV060268 Mus musculus
+	gb_est19:AA786530	7.00	680.68	120.06	282	AA786530 m3d04a1.r1 Aspergillus
-	gb_est46:AW386897	7.00	682.89	120.06	283	AW386897 CMO-PT0049-291299-140-
-	gb_est4:R89343	7.00	682.89	120.01	283	R89343 yq01g08.s1 Soares fetal
+	gb_est41:AV265974	7.00	687.30	119.98	285	AV265974 AV265974 RIKEN full-le
+	gb_est45:AV318192	7.00	689.51	119.96	286	AV318192 AV318192 RIKEN full-le
+	gb_est33:AV091932	7.00	691.71	119.96	287	AV091932 AV091932 Mus musculus
+	gb_est4:AI719435	7.00	691.96	119.93	287	AI719435 as64f02.x1 Barstead co
+	gb_est45:AW310643	7.00	693.92	119.88	288	AW310643 sg22f09.x1 Gm-c1024 Gl
+	gb_gss9:AQ020835	7.00	698.32	119.86	290	AQ020835 CIT-HSP-2302N17.TF CIT
+	gb_est34:AV166021	7.00	700.53	119.86	291	AV166021 AV166021 Mus musculus
+	gb_est2:F07466	7.00	700.53	119.86	291	F07466 HSC26G061 normalized inf
+	gb_est31:AV039250	7.00	700.53	119.86	291	AV039250 AV039250 Mus musculus
+	gb_est34:AV054410	7.00	700.53	119.86	291	AV054410 AV054410 Mus musculus
+	gb_est34:AV138888	7.00	702.73	119.84	291	AV138888 AV138888 Mus musculus
+	gb_est35:AV159347	7.00	704.93	119.81	292	AV159347 AV159347 Mus musculus
+	gb_est34:AB030410	7.00	707.13	119.79	293	AB030410 AB030410 Mus musculus
+	gb_est34:AV138102	7.00	713.73	119.71	294	AV138102 AV138102 Mus musculus
+	gb_est28:AI492221	7.00	718.12	119.67	297	AI492221 tg13a08.x1 NCI_CGAP_Ut
-	gb_est34:AV145356	7.00	720.32	119.64	299	AV145356 AV145356 Mus musculus
+	gb_est33:AV118645	7.00	720.32	119.64	300	AV118645 AV118645 Mus musculus
+	gb_gss3:B49634	7.00	720.32	119.64	300	B49634 CIT-HSP-202C2.TP CIT-HSP
+	gb_est16:C33655	7.00	720.32	119.64	300	C33655 C33655 Yuji Kohara unpub
+	gb_est16:C36892	7.00	720.32	119.64	300	C36892 C36892 Yuji Kohara unpub
+	gb_est16:C58723	7.00	722.51	119.62	300	C58723 C58723 Yuji Kohara unpub
+	gb_est36:AV176873	7.00	726.90	119.57	301	AV176873 AV176873 Yuji Kohara
+	gb_est36:AV178571	7.00	726.90	119.57	301	AV178571 AV178571 Yuji Kohara
+	gb_est36:AV181200	7.00	726.90	119.57	303	AV181200 AV181200 Yuji Kohara
+	gb_est36:AV182799	7.00	729.10	119.55	303	AV182799 AV182799 Yuji Kohara
+	gb_est36:AV183712	7.00	729.10	119.55	304	AV183712 AV183712 Yuji Kohara
-	gb_est40:AW151381	7.00	731.29	119.53	305	AW151381 xe75c06.x1 NCI_CGAP_Ut
+	gb_est1:Z44902	7.00	731.29	119.53	305	Z44902 UI-M-AJI-ahh-a-12-0-UI
+	gb_est2:T82315	7.00	735.67	119.48	307	T82315 AS306 Subtracted human
-	gb_est33:AV099448	7.00	735.67	119.48	307	AV099448 AV099448 Mus musculus
+	gb_gss1:FRO034589	7.00	737.86	119.46	308	AL030956 Fugu rubripes GSS sequ
-	gb_est19:AA790188	7.00	737.86	119.46	308	AA790188 vv82e09.r1 Stratagene
-	gb_est36:AI907152	7.00	740.06	119.43	309	AI907152 RC-BT133-110399-008 BT
-	gb_est1:T46996	7.00	742.25	119.41	310	T46996 yb12b06.s1 Stratagene pl
+	gb_est22:AA289218	7.00	744.43	119.39	311	AA289218 vc93c01.r1 Barstead MP
-	gb_est32:AV167325	7.00	744.43	119.39	311	AV167325 AV167325 Mus musculus
-	gb_est22:AI058788	7.00	746.62	119.36	312	AI058788 UI-R-C1-kt-e-01-0-UI.s
-	gb_est23:AW082426	7.00	746.62	119.36	312	AW082426 xb50b03.x1 NCI_CGAP_Es
-	gb_est46:AW322158	7.00	755.37	119.27	316	AW322158 uo52g09.x1 NCI_CGAP_Lu
+	gb_est22:AI021114	7.00	759.75	119.23	318	AI021114 ub01e01.r1 Soares_mamm
-	gb_est22:AA810415	7.00	766.30	119.16	321	AA810415 ob94a01.s1 NCI_CGAP_GC
-	gb_est27:AI399494	7.00	766.30	119.16	321	AI399494 NCSP6B6T3 Subtracted F
-	gb_est34:AV169605	7.00	766.30	119.16	321	AV169605 AV169605 Mus musculus
-	gb_est27:AI426127	7.00	772.85	119.07	324	AI426127 mj11g06.x1 Soares mous
+	gb_est5:D64557	7.00	775.03	119.07	325	D64557 CELK054FXR Yuji Kohara u
-	gb_est15:C25194	7.00	775.03	119.07	325	C25194 C25194 Rice shoot Oryza
-	gb_est22:AI020366	7.00	775.03	119.07	325	AI020366 ub27e01.r1 Soares_thym

Right block (ranks 325–373)

Rank	Pos	Tm		Strand	Oligo ID	Accession / Clone
325	775.03	119.07	7.00	+	gb_est27:AI438970	AI438970 tc89cll.x1 NCI_CGAP
325	775.03	119.07	7.00	-	gb_est35:AI848585	AI848585 UI-M-AM1-aga-d-05-0
326	777.21	119.05	7.00	-	gb_est1:D33751	D33751 CELK035E6R Yuji Kohar
326	777.21	119.05	7.00	+	gb_est2:F07259	F07259 HSC20D121 normalized
326	777.21	119.05	7.00	+	gb_est36:AI907735	AI907735 RC-BT152-230399-017
328	781.57	119.01	7.00	+	gb_est39:AV216215	AV216215 AV216215 RIKEN full
328	781.57	119.01	7.00	-	gb_est27:AA969466	AA969466 oo81d01.s1 NCI_CGAP
329	783.75	118.98	7.00	-	gb_est27:AI401117	AI401117 tg26b03.x1 NCI_CGAP
330	785.92	118.96	7.00	+	gb_est24:AI181532	AI181532 uc58h09.r1 Soares_t
331	788.14	118.94	7.00	+	gb_est36:AI895626	AI895626 EST265069 tomato ca
332	790.28	118.92	7.00	+	gb_est34:AI797253	AI797253 we86d07.x1 Soares_N
332	790.28	118.92	7.00	-	gb_est5:D71171	D71171 CELK063BYR Yuji Kohar
333	792.46	118.90	7.00	-	gb_est10:AA154679	AA154679 mt65f09.r1 Soares m
332	794.63	118.92	7.00	-	gb_est22:AI015916	AI015916 ov26b02.x1 Soares_t
334	794.63	118.88	7.00	-	gb_est11:AA237294	AA237294 mw96e10.r1 Soares m
334	794.63	118.88	7.00	-	gb_est24:AI208107	AI208107 qg51h09.x1 Soares_t
334	796.81	118.86	7.00	-	gb_est38:AW085072	AW085072 xe5c03.x1 Soares_N
335	796.81	118.86	7.00	+	gb_gss3:B31970	B31970 HS-1014-B1-A07-MF.ab1
335	801.16	118.81	7.00	+	gb_est25:AI277040	AI277040 q142c06.x1 NCI_CGAP
337	803.33	118.79	7.00	-	gb_est25:AI538685	AI538685 tp66e09.x1 NCI_CGAP
338	803.33	118.79	7.00	+	gb_est7:W44188	W44188 mc69f07.r1 Soares mou
338	805.51	118.77	7.00	-	gb_est5:D66608	D66608 CELK119D2R Yuji Kohar
339	807.68	118.75	7.00	+	gb_est29:AI600732	AI600732 486067H06.x1 486-
340	812.00	118.71	7.00	-	gb_est25:AI307235	AI307235 tb18c07.x1 NCI_CGAP
340	816.37	118.67	7.00	-	gb_gss4:AQ676796	AQ676796 HS_5496_B2_E05_SP6E
342	816.37	118.67	7.00	-	gb_est11:AA233593	AA233593 zr30f11.r1 Stratage
344	818.54	118.65	7.00	+	gb_est2:R03789	R03789 pk12b02.s1 Kuwabara M
344	818.54	118.65	7.00	-	gb_est32:AI732271	AI732271 yi67c12.x5 Soares b
345	820.71	118.63	7.00	-	gb_est12:AA297774	AA297774 EST113325 Infant br
345	822.87	118.61	7.00	-	gb_est45:AW237469	AW237469 xm72c11.x1 NCI_CGAP
346	822.87	118.61	7.00	-	gb_est47:AW412671	AW412671 uq48b08.x1 NCI_CGAP
347	825.04	118.58	7.00	+	gb_est24:AI207050	AI207050 qf80c12.x1 Soares_f
347	827.21	118.56	7.00	+	gb_est28:AI511601	AI511601 SWoVAFCAP29E10SK On
348	827.21	118.56	7.00	+	gb_est21:AA965725	AA965725 o4e08a1.r1 Aspergil
349	829.38	118.54	7.00	+	gb_est29:AI570545	AI570545 ln77h08.x1 NCI_CGAP
349	829.38	118.54	7.00	-	gb_gss12:AQ315865	AQ315865 RPCI11-1L3.TU RPCI-
349	831.55	118.52	7.00	-	gb_est21:AA670593	AA670593 v106g06.r2 Soares_m
350	831.55	118.52	7.00	+	gb_est37:AI944158	AI944158 614036E02.x1 614-
350	835.88	118.48	7.00	-	gb_est11:AA288304	AA288304 mr51e10.r1 Life Tec
351	835.88	118.48	7.00	-	gb_est26:AI390287	AI390287 mw96e10.y1 Soares m
351	835.88	118.48	7.00	+	gb_est5:N35442	N35442 yx90b10.r1 Soares mel
353	838.04	118.46	7.00	+	gb_est20:AA883386	AA883386 a156b07.s1 Soares_N
353	838.04	118.46	7.00	+	gb_est30:AI639111	AI639111 rx04444s Rat mixed-
353	838.04	118.46	7.00	+	gb_gss15:AQ555076	AQ555076 RPCI-11-420P19.TV R
353	838.04	118.46	7.00	+	gb_est1:Z44009	Z44009 HSCIQE091 normalized
354	838.04	118.46	7.00	+	gb_est21:AA921851	AA921851 om44d01.s1 Soares m
354	840.21	118.44	7.00	+	gb_est38:AW040197	AW040197 EST282696 tomato m1
354	842.37	118.42	7.00	+	gb_est38:AW040206	AW040206 EST282705 tomato m1
354	842.37	118.42	7.00	+	gb_est38:AW091609	AW091609 EST284885 tomato aort
354	846.70	118.38	7.00	-	gb_est42:AW161231	AW161231 au70b08.y1 Schneide
355	846.70	118.38	7.00	+	gb_est19:AI749323	AI749323 ny12c01.s1 NCI_CGAP
356	848.86	118.36	7.00	-	gb_est9:AW452349	AW452349 UI-H-BI3-alr-c-07-0
356	848.86	118.36	7.00	+	gb_est3:R55128	R55128 yg87h06.s1 Soares inf
358	851.03	118.34	7.00	+	gb_gss10:AQ095511	AQ095511 HS_3012_B1_D10_MR C
358	851.03	118.34	7.00	-	gb_est22:AI044716	AI044716 ot9g11.x1 Soares mous
359	851.03	118.34	7.00	+	gb_est31:AI669405	AI669405 ty32b06.x1 NCI_CGAP
359	851.03	118.34	7.00	+	gb_est6:D69938	D69938 CELK094AZF Yuji Kohar
359	853.19	118.32	7.00	+	gb_est5:D79542	D79542 HUM276F05B Human aort
360	855.35	118.30	7.00	-	gb_est16:C60758	C60758 C60758 Yuji Kohara un
360	857.51	118.28	7.00	-	gb_est17:C71576	C71576 C71576 Yuji Kohara un
360	859.67	118.26	7.00	+	gb_est2:D46377	D46377 RICS11008A Rice green
361	859.67	118.26	7.00	+	gb_est16:AA590774	AA590774 vm21b03.r1 Knowles
362	861.83	118.24	7.00	+	gb_est38:AA279481	AA279481 zs88h04.s1 Soares m1
363	863.99	118.22	7.00	-	gb_est38:AW048477	AW048477 UI-M-BH1-alj-c-03-0
364	863.99	118.22	7.00	-	gb_gss5:AQ770183	AQ770183 HS_2260_B1_C04_T7C
364	872.62	118.15	7.00	-	gb_gss10:AQ086897	AQ086897 HS_2200_A2_H07_MR C
365	874.78	118.13	7.00	+	gb_est21:AA916348	AA916348 oh80e05.s1 NCI_CGAP
366	876.93	118.11	7.00	+	gb_gss5:AQ804593	AQ804593 HS_2263_B1_A11_T7C
366	879.09	118.09	7.00	+	gb_gss11:AQ209669	AQ209669 HS_1238_A2_F09_MR C
370	879.09	118.09	7.00	-	gb_est47:AW416805	AW416805 52475 MARC 1P1G Sus
371			7.00	-	gb_est29:AI571822	AI571822 tn21f10.x1 NCI_CGAP
372			7.00	-	gb_est28:AI499779	AI499779 tm92d10.x1 NCI_CGAP
373			7.00	-	gb_est38:AW045820	AW045820 UI-M-BH1-aky-a-12-0

Left listing

gb_ID	str	7.00	score1	score2	rank	GenBank accession and description
gb_est17:C65099	-	7.00	118.07	881.24	374	C65099 C65099 Yuji Kohara unpub
gb_est25:AI309404	+	7.00	118.05	883.40	375	AI309404 tb37c09.x1 NCI_CGAP_HS
gb_est8:AA039025	+	7.00	118.03	885.55	376	AA039025 mi99e04.r1 Soares mous
gb_est16:C44658	+	7.00	118.03	885.55	376	C44658 C44658 Yuji Kohara unpub
gb_est17:C71614	+	7.00	118.03	885.55	376	C71614 C71614 Yuji Kohara unpub
gb_est19:AA747504	+	7.00	118.03	887.71	377	AA747504 nx77c03.s1 NCI_CGAP_Ew
gb_est10:AQ099298	+	7.00	118.01	889.86	378	AQ099298 HS_3042_B2_A01_MF CIT
gb_est1:D34021	-	7.00	117.99	889.86	378	D34021 CELK039G2R Yuji Kohara
gb_est24:AI166137	+	7.00	117.99	889.86	378	AI166137 B008p30u Hybrid aspen
gb_est29:AI566957	-	7.00	117.99	889.86	378	AI566957 tn25c09.x1 NCI_CGAP_Br
gb_est36:AV201161	+	7.00	117.99	896.32	381	AV201161 AV201161 Yuji Kohara u
gb_est28:AI490820	-	7.00	117.94	896.32	382	AI490820 EST241529 tomato shoot
gb_est11:AQ522581	-	7.00	117.92	898.47	382	AQ522581 HS_5240_A2_F05_T7A RPC
gb_est18:AA666029	-	7.00	117.92	898.47	382	AA666029 ag44b12.s1 Jia bone ma
gb_est35:AI852708	-	7.00	117.92	898.47	383	AI852708 UI-M-BH0-aj1-e-11-0-UI
gb_gss9:AQ031405	-	7.00	117.92	898.47	383	AQ031405 HS_2222_A2_E07_MF CIT
gb_est6:W11319	-	7.00	117.90	900.62	383	W11319 ma78c05.r1 Soares mouse
gb_est16:AA561354	+	7.00	117.86	904.92	385	AA561354 UI-18e08.r1 Stratagene
gb_est8:C10522	-	7.00	117.85	907.07	386	C10522 C10522 Yuji Kohara unpu
gb_est9:AA107760	-	7.00	117.85	907.07	386	AA107760 mo48g03.r1 Life Tech m
gb_est29:AI569458	-	7.00	117.83	909.22	387	AI569458 nf88d04.s1 NCI_CGAP_Ut
gb_est37:AI973249	-	7.00	117.83	909.22	387	AI973249 wr53f12.x1 NCI_CGAP_Ki
gb_est44:AW214739	+	7.00	117.79	913.52	389	AW214739 uo99c09.y1 NCI_CGAP_Lu
gb_est19:AA768437	-	7.00	117.79	913.52	389	AA768437 ob22f07.s1 NCI_CGAP_Ki
gb_est32:AI715296	-	7.00	117.79	913.52	389	AI715296 UI-R-Y0-abp-f-12-0-UI
gb_gss3:B33820	-	7.00	117.79	913.52	389	B33820 HS-1023-A2-E04-MF.abi CI
gb_gss3:B42625	+	7.00	117.79	913.52	389	B42625 HS-1056-A1-G08-MF.abi CI
gb_gss12:AQ312357	-	7.00	117.75	917.81	391	AQ312357 RPCI11-112M18.TJ RPCI-
gb_gss11:AA222474	+	7.00	117.74	919.96	392	AA222474 mw20e02.r1 Soares mous
gb_est39:AW123302	+	7.00	117.74	919.96	392	AW123302 UI-M-BH2.1-apg-h-11-0-
gb_est3:R65616	-	7.00	117.72	922.10	393	R65616 yi26b02.r1 Soares placen
gb_est44:AW206298	-	7.00	117.70	924.25	394	AW206298 UI-H-BI1-afe-g-12-0-UI
gb_est14:AQ507837	+	7.00	117.70	924.25	394	AQ507837 RPCI11-278D14.TV RPCI
gb_est8:W89921	+	7.00	117.68	926.39	395	W89921 mf77c09.r1 Soares mouse
gb_est11:AI596323	-	7.00	117.66	928.54	396	AI596323 ve24b03.y1 Soares mou
gb_est34:AI801623	-	7.00	117.66	928.54	396	AI801623 to91h03.x1 NCI_CGAP_Lu
gb_gss3:B35674	-	7.00	117.65	930.68	396	B35674 HS-1029-A2-H07-MR.abi CI
gb_est8:AQ187979	-	7.00	117.63	932.83	398	AQ187979 HS_3142_B2_E07_MR CIT
gb_est27:AI423456	-	7.00	117.61	934.97	399	AI423456 tf72e10.x1 NCI_CGAP_Br
gb_est11:AA989124	-	7.00	117.59	937.11	400	AA989124 or75h04.s1 NCI_CGAP_Lu
gb_est1:T00533	-	7.00	117.57	939.26	401	T00533 WEST01254 Early embryo,
gb_est15:AA515027	-	7.00	117.57	939.26	401	AA515027 nf64h06.s1 NCI_CGAP_Co
gb_gss7:AQ903740	-	7.00	117.57	939.26	401	AQ903740 GSSTc04781 Trypanosome
gb_est16:AQ101952	-	7.00	117.57	939.26	401	AQ101952 HS_3017_B1_G12_T7 CIT
gb_est11:AQ186130	-	7.00	117.56	941.40	402	AQ186130 HS_3077_A1_B11_MR CIT
gb_est2:T97145	-	7.00	117.56	941.40	402	T97145 ye50o10.r1 Soares fetal
gb_gss3:B67144	-	7.00	117.54	943.54	403	B67144 CIT-HSP-2017O24.TFB CIT
gb_est42:AW162091	+	7.00	117.52	945.68	404	AW162091 au98d07.y1 Schneider f
gb_est3:R24986	-	7.00	117.52	945.68	404	R24986 yg41h03.r1 Soares infant
gb_est24:AI242291	-	7.00	117.50	947.82	405	AI242291 qi16g02.x1 Soares NhHM
gb_est17:AA634492	-	7.00	117.50	947.82	405	AA634492 zu76d01.s1 Soares_test
gb_est34:AI795818	+	7.00	117.49	949.96	406	AI795818 RC0-HT0023-300699-001-
gb_est46:AW382288	-	7.00	117.49	949.96	407	AW382288 CM2-HT0326-291199-044
gb_gss3:B85591	+	7.00	117.47	952.10	408	B85591 RPCI11-20C11.TV RPCI-11
gb_est20:AA897470	-	7.00	117.45	954.24	408	AA897470 a148e04.s1 Soares_NFL_
gb_gss3:AL044310	+	7.00	117.45	954.24	408	AL044310 DKFZp434O0429_r1 434
gb_est45:AW245695	-	7.00	117.45	954.24	409	AW245695 2822917.5prime NIH_MGC
gb_est47:AW432376	+	7.00	117.43	956.38	409	AW432376 sh73b09.y1 Gm-c1015 Gl
gb_gss5:AQ822184	+	7.00	117.43	956.38	412	AQ822184 HS_5494_A1_E10_SP6E RF

Right listing

gb_ID	str	7.00	score1	score2	rank	GenBank accession and description
gb_gss12:AQ293177		7.00	117.38	962.79	412	AQ293177 HS_2201_B2_H04.MF C
gb_gss13:AQ422850		7.00	117.38	962.79	412	AQ422850 CITBI-E1-2577C12.TF
gb_est15:AA479063		7.00	117.35	967.07	414	AA479063 zv18h01.s1 Soares_N
gb_est16:AI361578		7.00	117.33	967.07	414	AI361578 qy48e11.x1 NCI_CGAP
gb_est45:AW244976		7.00	117.31	969.20	415	AW244976 687017H02.y1 687-
gb_est22:AI074659		7.00	117.31	971.34	416	AI074659 ox82g07.s1 Soares_N
gb_est23:AA099298		7.00	117.29	973.48	417	AA099298 ne38h08.s1 NCI_CGAP
gb_est22:AI003949		7.00	117.29	973.48	417	AI003949 ot94b06.x1 Soares_t
gb_est23:AI155914		7.00	117.29	973.48	417	AI155914 ue04a08.r1 Soares_N
gb_est24:AA398211		7.00	117.28	975.61	418	AA398211 zt59c02.s1 Soares_t
gb_est28:AI502971		7.00	117.28	975.61	418	AI502971 vm69c05.x1 Knowles
gb_est15:AA456702		7.00	117.26	977.74	419	AA456702 aa13f11.r1 Soares adu
gb_est15:H51345		7.00	117.24	979.88	420	H51345 yj67c12.y5 Soares b
gb_est35:AQ630326		7.00	117.24	979.88	420	AQ630326 RPCI-11-469J24.TJ R
gb_est35:AA935182		7.00	117.24	982.01	421	AA935182 oh33d07.s1 NCI_CGAP
gb_est24:AU029915		7.00	117.23	982.01	421	AU029915 AU029915 Rice cDNA
gb_est26:AA900894		7.00	117.23	982.01	421	AA900894 UI-R-E0-dc-a-10-0-U
gb_est16:AI566161		7.00	117.23	982.01	421	AI566161 tg69bl1.x1 NCI_CGAP
gb_est32:AI760984		7.00	117.23	982.01	421	AI760984 wh97a08.x1 NCI_CGAP
gb_est45:AW258986		7.00	117.23	982.01	421	AW258986 um76a04.y1 Sugano m
gb_est29:AI219894		7.00	117.21	982.01	421	AI219894 HS_3252_A2_G11.MR C
gb_est30:AI662710		7.00	117.19	984.15	422	AI662710 va86h11.x1 Soares m
gb_gss5:AQ807397		7.00	117.18	984.15	422	AQ807397 HS_3177_A2_D11.MR C
gb_gss11:AI307402		7.00	117.18	986.28	423	AI307402 tb26h01.x1 NCI_CGAP
gb_gss13:AQ371431		7.00	117.16	988.41	424	AQ371431 HS_5043_A1_F04_SP6E
gb_est30:AQ402925		7.00	117.14	988.41	424	AQ402925 HS_5066_A2_E12_SP6E
gb_gss14:AQ499762		7.00	117.14	988.41	424	AQ499762 HS_5204_A2_B09_SP6E
gb_est20:AA870578		7.00	117.14	990.54	425	AA870578 vq23d10.r1 Barstead
gb_est37:AW008196		7.00	117.13	992.67	426	AW008196 wv51d02.x1 NCI_CGAP
gb_gss5:AQ675929		7.00	117.11	992.67	426	AQ675929 HS_5159_A1_G10_T7C
gb_gss5:AQ813209		7.00	117.08	992.67	426	AQ813209 HS_5243_B1_E11_SP6E
gb_gss11:AQ403791		7.00	117.08	994.80	427	AQ403791 HS_5049_A1_C01_SP6E
gb_est37:AI943657		7.00	117.06	996.94	428	AI943657 618042H12.x1 618-
gb_est14:AA406992		7.00	117.04	1.0e+03	430	AA406992 EST01998 Mouse 7.5
gb_est16:AA583432		7.00	117.08	1.0e+03	430	AA583432 nn38a08.s1 NCI_CGAP
gb_est23:AA836857		7.00	117.08	1.0e+03	430	AA836857 of35a05.s1 NCI_CGAP
gb_est6:D79572		7.00	117.06	1.0e+03	431	D79572 HUM281G04B Human aort
gb_est11:AQ208520		7.00	117.04	1.0e+03	431	AQ208520 HS_3237_A1_A04_MR C
gb_est18:AA694431		7.00	117.04	1.0e+03	432	AA694431 ah22b05.s1 Soares_p
gb_est23:AI082401		7.00	117.04	1.0e+03	432	AI082401 ov1c10.x1 NCI_CGAP
gb_est27:AI421702		7.00	117.04	1.0e+03	432	AI421702 tf54d09.x1 NCI_CGAP
gb_est39:AW105955		7.00	117.04	1.0e+03	432	AW105955 um20c12.y1 Sugano m
gb_est39:AW126197		7.00	117.04	1.0e+03	432	AW126197 N100042e rootphos(-
gb_gss4:AA704266		7.00	117.04	1.0e+03	432	AA704266 HS_5475_B2_H10_SP6E
gb_est19:AA760813		7.00	117.03	1.0e+03	433	AA760813 nz10f05.s1 NCI_CGAP
gb_est19:AA776881		7.00	117.03	1.0e+03	433	AA776881 ac40a08.s1 Stratage
gb_est28:AA822155		7.00	117.03	1.0e+03	434	AA822155 vp35e08.r1 Barstead
gb_est28:AI506159		7.00	117.01	1.0e+03	435	AI506159 vp35e08.x1 Barstead
gb_est5:N26282		7.00	116.99	1.0e+03	435	N26282 yw62h10.s1 Soares_pla
gb_est10:AA157978		7.00	116.99	1.0e+03	435	AA157978 zo71h07.s1 Stratage
gb_est22:AA162230		7.00	116.99	1.0e+03	435	AA162230 mn44h01.r1 Bedding
gb_est22:AI004733		7.00	116.98	1.0e+03	435	AI004733 ou04e06.x1 Soares_f
gb_est28:C95514		7.00	116.96	1.0e+03	436	C95514 C95514 Citrus unshiu
gb_est32:AI762505		7.00	116.96	1.0e+03	437	AI762505 wh91d02.x1 NCI_CGAP
gb_gss4:AA686865		7.00	116.93	1.0e+03	439	AA686865 nbxb0073O15r CUGI R
gb_gss9:B95967		7.00	116.93	1.0e+03	439	B95967 T30K5TR TAMU Arabidop
gb_est4:D57179		7.00	116.91	1.0e+03	440	D57179 HUM274D04B Clontech h
gb_est42:AI161562		7.00	116.90	1.0e+03	441	AI161562 au82h03.y1 Schneide
gb_est21:AI245725		7.00	116.90	1.0e+03	441	AI245725 qk26e03.x1 NCI_CGAP
gb_est25:AA009744		7.00	116.88	1.0e+03	442	AA009744 ze82g01.s1 Soares_f
gb_gss11:AQ255415		7.00	116.88	1.0e+03	442	AQ255415 mgxb0012P10r CUGI R
gb_est22:AA154861		7.00	116.86	1.0e+03	443	AA154861 mq6if03.r1 Soares_t
gb_est24:AI016465		7.00	116.86	1.0e+03	443	AI016465 ot78h07.s1 Soares_t
gb_gss9:AI190323		7.00	116.86	1.0e+03	443	AI190323 qd57h06.x1 Soares_t
gb_gss2:AQ077025		7.00	116.85	1.0e+03	443	AQ077025 CIT-HSP-2367A18.TF
gb_est14:R08375		7.00	116.85	1.0e+03	444	R08375 yf18c09.s1 Soares fet
gb_est45:AA417921		7.00	116.85	1.0e+03	444	AA417921 zv94b03.s1 Soares_N
gb_gss5:AW284387	+	7.00	116.85	1.0e+03	444	AW284387 LG1_275_H08.g1_A002

Left listing (rows 444–468):

flag	query	#	E	bits	q	subject
-	gb_gss3:B52705	444	1.0e+03	116.85	7.00	B52705 CIT-HSP-206I6.TR CIT-HS
.	gb_est9:AA19938	445	1.0e+03	116.83	7.00	AA119938 mp84d03.r1 Soares_thym
-	gb_est10:AA135185	447	1.0e+03	116.80	7.00	AA135185 zo27a05.s1 Stratagene
+	gb_est35:AI875271	447	1.0e+03	116.80	7.00	AI875271 uk30b04.y1 Sugano mous
+	gb_est37:AI999418	447	1.0e+03	116.80	7.00	AI999418 701555815 A. thaliana,
+	gb_est38:AW046039	447	1.0e+03	116.80	7.00	AW046039 UI-M-BHI-akv-b-05-0-UI
-	gb_est38:AW073711	447	1.0e+03	116.80	7.00	AW073711 xb01h10.x1 NCI_CGAP_GU
+	gb_est42:AW160789	447	1.0e+03	116.80	7.00	AW160789 au66b05.y1 Schneider A
+	gb_est45:AW291316	447	1.0e+03	116.80	7.00	AW291316 UI-H-B12-agg-c-12-0-UI
-	gb_est19:AA760063	448	1.0e+03	116.78	7.00	AA760063 vv69a09.r1 Stratagene
+	gb_est26:AI366778	448	1.0e+03	116.78	7.00	AI366778 qq41d06.x1 Soares_NhHM
-	gb_est35:AI816334	448	1.0e+03	116.78	7.00	AI816334 au46c01.y1 Schneider f
-	gb_gss11:AQ200590	448	1.0e+03	116.78	7.00	AQ200590 RPCI11-43H7.TK RPCI-11
-	gb_est1:D26732	449	1.0e+03	116.77	7.00	D26732 CELK010GYF Yuji Kohara u
+	gb_est23:AI140962	449	1.0e+03	116.77	7.00	AI140962 qf24f01.x1 NCI_CGAP_Br
+	gb_est28:AI288600	449	1.0e+03	116.77	7.00	AI288600 q182g12.x1 Soares_NhHM
+	gb_est36:AI931292	449	1.0e+03	116.77	7.00	AI931292 u167c10.y1 Sugano mous
+	gb_gss8:AQ984707	449	1.0e+03	116.77	7.00	AQ984707 RPCI-23-266O15.TV RPCI
-	gb_est10:AQ122354	449	1.0e+03	116.77	7.00	AQ122354 HS_3087_A1_D11_MR CIT
-	gb_est8:AA16114	450	1.0e+03	116.75	7.00	AA016114 ze3if08.r1 Soares reti
+	gb_est14:AA418172	450	1.0e+03	116.75	7.00	AA418172 zv94f03.r1 Soares_NhHM
+	gb_est18:AA768438	450	1.0e+03	116.75	7.00	AA768438 ob22f08.s1 NCI_CGAP_Ki
+	gb_gss14:AQ448553	450	1.0e+03	116.75	7.00	AQ448553 mgxb0021A13f CUGI Rice
+	gb_est2:D46560	451	1.0e+03	116.74	7.00	D46560 RICS11315A Rice green sh
+	gb_est3:R25343	451	1.0e+03	116.74	7.00	R25343 yh42d05.r1 Soares placen
+	gb_est11:AA210600	451	1.0e+03	116.74	7.00	AA210600 mu71g01.r1 Soares mous
+	gb_est12:AA324494	451	1.0e+03	116.74	7.00	AA324494 EST27613 Cerebellum II
+	gb_est18:AA684357	451	1.0e+03	116.74	7.00	AA684357 vm69d10.s1 Knowles Sol
+	gb_est38:AW029472	451	1.0e+03	116.74	7.00	AW029472 wx12b09.x1 NCI_CGAP_Ga
+	gb_gss5:AQ805512	451	1.0e+03	116.74	7.00	AQ805512 HS_3094_A2_H07_MR CIT
+	gb_est29:AI557599	452	1.0e+03	116.72	7.00	AI557599 pt2.1-1.G05.r tumor2 H
+	gb_est5:AQ769083	453	1.0e+03	116.70	7.00	AQ769083 HS_3155_A2_B11_MR CIT
-	gb_est10:AI163527	453	1.1e+03	116.70	7.00	AI163527 ms22g12.r1 Stratagene
+	gb_est25:AI263516	453	1.1e+03	116.70	7.00	AI263516 q120b03.x1 Soares_NhHM
+	gb_est34:AI801537	453	1.1e+03	116.70	7.00	AI801537 to90f12.x1 NCI_CGAP_Ga
+	gb_est8:AA034841	454	1.1e+03	116.69	7.00	AA034841 mi52f09.r1 Soares mous
+	em_est30:HSM000002	454	1.1e+03	116.69	7.00	AL035711 Homo sapiens mRNA; EST
+	gb_gss1:FR002852	454	1.1e+03	116.69	7.00	AL024824 Fugu rubripes GSS sequ
+	gb_gss7:AQ364491	454	1.1e+03	116.69	7.00	AQ364491 nbxb0061I13r CUGI Rice
+	gb_est14:AW040342	455	1.1e+03	116.67	7.00	AW040342 EST283206 tomato mixed
+	gb_est46:AW402696	455	1.1e+03	116.67	7.00	AW402696 UI-HF-BKO-aav-g-06-0-U
+	gb_est4:AQ696503	455	1.1e+03	116.67	7.00	AQ696503 HS_5518_A2_H09_SP6E RE
+	gb_est29:AI775789	455	1.1e+03	116.67	7.00	AI775789 HS_2006_A2_E03_MR CIT
+	gb_est7:W78390	456	1.1e+03	116.66	7.00	W78390 me78c07.r1 Soares mouse
+	gb_est45:AW308630	456	1.1e+03	116.66	7.00	AW308630 SWOv3MCAM37C11SK Onchc
+	gb_gss3:B75677	456	1.1e+03	116.66	7.00	B75677 RPCI11-11P13.TP RPCI-11
+	gb_gss6:AQ861662	456	1.1e+03	116.64	7.00	AQ861662 nbeb0017C12r CUGI Rice
+	gb_gss13:AQ371205	457	1.1e+03	116.63	7.00	AQ371205 HS_5048_A1_A08_SP6E RE
+	gb_est15:W71069	458	1.1e+03	116.63	7.00	W71069 me31e08.r1 Soares mouse
+	gb_est24:AI211467	458	1.1e+03	116.63	7.00	AI211467 p0d10a1.r1 Aspergillus
+	gb_est30:AI308941	458	1.1e+03	116.63	7.00	AI308941 qo65d06.x1 NCI_CGAP_Co
+	gb_est30:AI644639	458	1.1e+03	116.63	7.00	AI644639 vw79a07.x1 Stratagene
+	gb_est16:AA556009	460	1.1e+03	116.59	7.00	AA556009 vk49e06.r1 Stratagene
+	gb_est18:AI509975	461	1.1e+03	116.58	7.00	AI509975 m115g06.r1 Soares mous
+	gb_est32:AI722851	461	1.1e+03	116.58	7.00	AI722851 fc32e03.y1 Zebrafish W
+	gb_est46:AW329937	461	1.1e+03	116.58	7.00	AW329937 TENU4744 T.cruzi epima
+	gb_gss5:AQ771137	461	1.1e+03	116.58	7.00	AQ771137 HS_5379_B2_G10_SP6E RE
+	gb_gss6:AQ862698	461	1.1e+03	116.58	7.00	AQ862698 nbeb0019I22f CUGI Rice
.	gb_est17:C72082	462	1.1e+03	116.56	7.00	C72082 C72082 Rice panicle at f
+	gb_est10:AA163104	463	1.1e+03	116.55	7.00	AA163104 mr93c11.r1 Stratagene
+	gb_est28:AI505513	463	1.1e+03	116.55	7.00	AI505513 vn37h10.x1 Stratagene
+	gb_gss9:AQ060496	463	1.1e+03	116.55	7.00	AQ060496 CIT-HSP-2350K1.TR CIT-
.	gb_est19:AA757951	464	1.1e+03	116.53	7.00	AA757951 vw57a09.r1 Soares_mamm
+	gb_est38:AW048858	465	1.1e+03	116.52	7.00	AW048858 UI-M-BHI-amf-b-12-0-UI
+	gb_est3:R28109	465	1.1e+03	116.49	7.00	R28109 yh58c06.r1 Soares placen
+	gb_est15:AA510271	467	1.1e+03	116.49	7.00	AA510271 vh58d04.r1 Soares_mamm
+	gb_est31:AI692133	467	1.1e+03	116.49	7.00	AI692133 SWOv3MCAM28E06SK Onchc
+	gb_est38:AW047580	467	1.1e+03	116.49	7.00	AW047580 UI-M-BHI-akm-b-01-0-UI
+	gb_est38:AW052797	467	1.1e+03	116.49	7.00	AW052797 614036E02.y1 614 - roc
+	gb_est45:AW259875	468	1.1e+03	116.47	7.00	AW259875 um79f05.y1 Sugano thym
+	gb_est9:AA125259	468	1.1e+03	116.47	7.00	AA125259 mp71d03.r1 Soares_thym

Right listing (rows 469–498):

flag	query	#	E	bits	q	subject
-	gb_est23:AI147693	469	1.1e+03	116.46	7.00	AI147693 qb43a10.x1 NCI_CGAP
-	gb_est26:AI347038	469	1.1e+03	116.46	7.00	AI347038 qp54h11.x1 NCI_CGAP
.	gb_gss13:AQ401242	469	1.1e+03	116.44	7.00	AQ401242 HS_5066_A1_G01_SP6E
.	gb_est20:F22874	470	1.1e+03	116.44	7.00	F22874 SSC16S08 Porcine smal
+	gb_est46:AV403482	470	1.1e+03	116.43	7.00	AV403482 AV403482 Bombyx mor
-	gb_est8:AA036083	471	1.1e+03	116.43	7.00	AA036083 mi73h02.r1 Soares m
-	gb_est23:AA869032	471	1.1e+03	116.43	7.00	AA869032 vq25b10.r1 Barstead
-	gb_est28:AI098344	471	1.1e+03	116.43	7.00	AI098344 vg84b12.r1 Barstead
+	gb_est39:AW130018	471	1.1e+03	116.41	7.00	AW130018 xf26h03.x1 NCI_CGAP
-	gb_est8:W89543	472	1.1e+03	116.41	7.00	W89543 mf71c10.r1 Soares mou
+	gb_est27:AI432456	472	1.1e+03	116.41	7.00	AI432456 tg73g06.x1 Soares_N
	gb_est44:AJ273531	472	1.1e+03	116.41	7.00	AJ273531 Metarhiziu
	gb_est46:AW318537	472	1.1e+03	116.40	7.00	AW318537 um99a02.y1 Sugano m
	gb_est31:AI685261	473	1.1e+03	116.38	7.00	AI685261 wa75e10.x1 Soares_N
+	gb_gss7:AQ972473	473	1.1e+03	116.38	7.00	AQ972473 RPCI-23-312O21.TJ R
	gb_est15:AA528216	474	1.1e+03	116.37	7.00	AA528216 nj16h06.s1 NCI_CGAP
	gb_gss5:AQ011421	474	1.1e+03	116.37	7.00	AQ11421 HS_5226_B1_D11_SP6E
	gb_gss7:AQ941433	474	1.1e+03	116.35	7.00	AQ941433 Sheared DNA-44M18.T
	gb_est2:N98437	475	1.1e+03	116.37	7.00	N98437 zb77g09.s1 Soares_sen
	gb_est24:AI211328	475	1.1e+03	116.37	7.00	AI211328 o6d01a1.r1 Aspergil
	gb_est29:AI554501	475	1.1e+03	116.37	7.00	AI554501 tq22c09.x1 NCI_CGAP
	gb_est7:W87316	476	1.1e+03	116.35	7.00	W87316 zh64d05.r1 Soares_fet
	gb_est17:C78791	476	1.1e+03	116.35	7.00	C78791 C78791 Mouse 3.5-dpc
	gb_est25:AI290167	476	1.1e+03	116.35	7.00	AI290167 q18d12.x1 Soares_N
	gb_gss13:AQ389481	476	1.1e+03	116.35	7.00	AQ389481 RPCI11-153H22.TV RP
	gb_gss14:AQ516506	476	1.1e+03	116.35	7.00	AQ516506 HS_5072_B2_A10_T7A
	gb_est4:H23275	477	1.1e+03	116.34	7.00	H23275 ym57f04.s1 Soares_inf
	gb_est30:AA847787	478	1.1e+03	116.32	7.00	AA847787 oe99h10.s1 NCI_CGAP
	gb_est38:AW058087	478	1.1e+03	116.32	7.00	AW058087 wv93b06.x1 Soares_t
	gb_est5:N26295	479	1.1e+03	116.31	7.00	N26295 yw62f10.s1 Soares_pla
	gb_est8:AA030424	479	1.1e+03	116.31	7.00	AA030424 mi28g04.r1 Soares m
	gb_est23:AI091181	479	1.1e+03	116.31	7.00	AI091181 oo18d07.x1 Soares m
	gb_est45:AW259944	479	1.1e+03	116.31	7.00	AW259944 um79g01.y1 Sugano m
	gb_gss11:AQ227059	479	1.1e+03	116.31	7.00	AQ227059 HS_2024_B1_A10_MR C
	gb_est4:H26363	481	1.1e+03	116.28	7.00	H26363 yl44c07.s1 Soares bre
	gb_est17:AA632915	481	1.1e+03	116.28	7.00	AA632915 nm14g01.s1 NCI_CGAP
	gb_est45:AI459800	481	1.1e+03	116.28	7.00	AI459800 ap17g02.x1 Schiller
	gb_est8:AW275342	482	1.1e+03	116.26	7.00	AW275342 xm79c07.x1 NCI_CGAP
	gb_est8:AA021675	482	1.1e+03	116.26	7.00	AA021675 mh82g07.r1 Soares m
	gb_gss7:AQ364250	482	1.1e+03	116.26	7.00	AQ364250 nbxb0060P21f CUGI R
	gb_gss15:AQ547651	483	1.1e+03	116.25	7.00	AQ547651 RPCI-11-382C6.TJ RP
	gb_est8:W88219	483	1.1e+03	116.25	7.00	W88219 mf70c05.r1 Soares mou
	gb_est27:AI430352	484	1.1e+03	116.23	7.00	AI430352 mff7c10.y1 Soares m
	gb_est21:AA982967	484	1.1e+03	116.23	7.00	AA982967 ua23h04.r1 Soares m
	gb_est23:AI088652	484	1.1e+03	116.23	7.00	AI088652 qb14a07.x1 Soares_p
	gb_est27:AI909290	484	1.1e+03	116.20	7.00	AI909290 EST237582 Normalize
	gb_est37:AI960688	486	1.1e+03	116.20	7.00	AI960688 sc89a06.y1 Gm-c1019
	gb_est23:AI116958	486	1.1e+03	116.20	7.00	AI116958 ue26a07.y1 Sugano m
	gb_est29:AI593585	486	1.1e+03	116.19	7.00	AI593585 vp35e08.y1 Barstead
	gb_est36:AI931202	487	1.1e+03	116.19	7.00	AI931202 u157f08.y1 Sugano m
	gb_est16:AA551035	487	1.1e+03	116.17	7.00	AA551035 nk74b05.s1 NCI_CGAP
	gb_est22:AI057601	488	1.1e+03	116.17	7.00	AI057601 oy31e01.x1 Soares m
	gb_est10:AA171007	488	1.1e+03	116.16	7.00	AA171007 ms54d08.r1 Life Tec
	gb_est14:AA437783	489	1.1e+03	116.16	7.00	AA437783 ve33e01.r1 Ko mouse
	gb_est11:AA200128	490	1.1e+03	116.14	7.00	AA200128 mu10a05.r1 Soares m
	gb_est34:AI800943	491	1.1e+03	116.13	7.00	AI800943 wg06a05.x1 Soares_N
	gb_est21:AA936255	491	1.1e+03	116.12	7.00	AA936255 on75b04.s1 Soares_N
	gb_est21:AA929398	492	1.1e+03	116.10	7.00	AA929398 vz53g07.r1 Soares_t
	gb_est45:AW279746	493	1.1e+03	116.10	7.00	AW279746 fj44b12.x1 zebrafis
	gb_est15:AA522038	493	1.1e+03	116.09	7.00	AA522038 v108e11.r1 Barstead
	gb_est37:AI979623	494	1.1e+03	116.09	7.00	AI979623 614038C08.x2 614 -
	gb_gss16:AQ621499	494	1.1e+03	116.10	7.00	AQ621499 HS_3029_B3_F05_MF C
	gb_est27:AI422091	494	1.1e+03	116.09	7.00	AI422091 tf57g11.x1 NCI_CGAP
	gb_est34:AI802199	494	1.1e+03	116.09	7.00	AI802199 tx25d07.x1 NCI_CGAP
	gb_est37:AI979583	495	1.1e+03	116.07	7.00	AI979583 614036E02.x2 614 -
	gb_est45:AW289665	496	1.1e+03	116.06	7.00	AW289665 NXNV004A04F Nsf Xyl
	gb_gss16:AQ551927	497	1.1e+03	116.06	7.00	AQ551927 RPCI-11-42S09.TJ RP
	gb_gss16:AQ665932	497	1.1e+03	116.04	7.00	AQ665932 HS_5372_B2_A07_SP6E
	gb_gss14:AQ472930	498	1.1e+03	116.03	7.00	AQ472930 CITBI-E1-2589A8.TR
	gb_est45:AV393530	498	1.1e+03	116.03	7.00	AV393530 AV393530 chlamydomo
	gb_est45:AW249397	498	1.1e+03	116.03	7.00	AW249397 2819379.5prime NIH_
	gb_gss12:AQ275949	498	1.1e+03	116.03	7.00	AQ275949 CITBI-E1-2524I16.TR
	gb_gss14:AQ492751	498	1.1e+03	116.03	7.00	AQ492751 HS_5219_B2_A07_SP6E

Left block

flag	id	val	score		description	accession	#
-	gb_est23:AI091239	1.1e+03	116.01	7.00	oo26b01.x1 Soares_NSF	AI091239	499
+	gb_est28:AI545064	1.1e+03	116.01	7.00	fb70c05.y1 Zebrafish W	AI545064	499
-	gb_est38:AW071742	1.1e+03	116.01	7.00	w855a12.x1 NCI_CGAP_Br	AW071742	499
-	gb_est45:AW260594	1.1e+03	116.00	7.00	um83f11.y1 Sugano mous	AW260594	499
+	gb_est14:AA412292	1.1e+03	116.00	7.00	zu1o9o6.s1 Soares_test	AA412292	500
-	gb_est20:AA898726	1.1e+03	115.99	7.00	NCM6E10T7 Mycelial Neu	AA898726	500
-	gb_est15:AA524995	1.2e+03	115.99	7.00	nh32c02.s1 NCI_CGAP_A1	AA524995	501
-	gb_est23:AI098370	1.2e+03	115.99	7.00	vg84f11.r1 Barstead MF	AI098370	501
+	gb_est44:AV394978	1.2e+03	115.99	7.00	AV394978 Chlamydomonas	AV394978	501
-	gb_est44:AW222157	1.2e+03	115.99	7.00	EST298968 tomato fruit	AW222157	501
-	gb_gss16:AQ671354	1.2e+03	115.99	7.00	HS_5458_A2_C01_SP6E RE	AQ671354	501
-	gb_est36:AI891634	1.2e+03	115.94	7.00	u157f08.x1 Sugano mous	AI891634	504
+	gb_gss9:AQ015726	1.2e+03	115.93	7.00	CIT-HSP-2307I4.TR CIT	AQ015726	505
-	gb_gss12:AQ335831	1.2e+03	115.93	7.00	HS_5016_B1_G12_SP6E RE	AQ335831	505
-	gb_est18:AA672957	1.2e+03	115.90	7.00	vn48b03.r1 Barstead mc	AA672957	507
-	gb_est25:AI317173	1.2e+03	115.90	7.00	u133h09.y1 Sugano mous	AI317173	507
+	gb_est8:AA009462	1.2e+03	115.89	7.00	ze82g01.r1 Soares_feta	AA009462	508
-	gb_est8:AA057520	1.2e+03	115.89	7.00	zf56d01.s1 Soares_reti	AA057520	508
-	gb_est29:AI607620	1.2e+03	115.89	7.00	vk49e06.y1 Stratagene	AI607620	508
-	gb_est36:AI930148	1.2e+03	115.89	7.00	u161c08.y1 Sugano mous	AI930148	508
+	gb_est39:AW107071	1.2e+03	115.87	7.00	u192f02.y1 Sugano mous	AW107071	509
+	gb_est29:AI607665	1.2e+03	115.87	7.00	va86h11.y1 Soares mous	AI607665	509
-	gb_est37:AI994270	1.2e+03	115.87	7.00	701501945 A. thaliana,	AI994270	509
-	gb_gss6:AQ858570	1.2e+03	115.87	7.00	nbeb0002B11r CUGI Rice	AQ858570	509
-	gb_est11:AA271129	1.2e+03	115.86	7.00	va86h11.r1 Soares mous	AA271129	510
+	gb_gss1:CNS00YIE	1.2e+03	115.86	7.00	Arabidopsis thaliana g	CNS00YIE	510
-	gb_gss4:AQ687728	1.2e+03	115.86	7.00	nbxb0075F09r CUGI Rice	AQ687728	510
-	gb_gss10:AQ132220	1.2e+03	115.86	7.00	HS_3005_A1_E10_MF CIT	AQ132220	510
-	gb_est21:AA939273	1.2e+03	115.85	7.00	oq31b07.s1 NCI_CGAP_GC	AA939273	511
-	gb_gss3:B27068	1.2e+03	115.85	7.00	T27F7TF TAMU Arabidopsis	B27068	511
-	gb_est13:AQ369533	1.2e+03	115.85	7.00	HS_5033_B2_B08_T7 RPCI	AQ369533	511
-	gb_gss16:AQ620640	1.2e+03	115.85	7.00	HS_5183_B2_C04_SP6E RE	AQ620640	511
+	gb_est35:AI854072	1.2e+03	115.83	7.00	UI-M-BHO-ajm-h-01-0-UI	AI854072	512
-	gb_est18:AW069290	1.2e+03	115.80	7.00	w168c06.x1 NCI_CGAP_Br	AW069290	514
+	gb_est38:AW206294	1.2e+03	115.80	7.00	EST285469 tomato mixed	AW206294	514
+	gb_est44:AW206294	1.2e+03	115.80	7.00	UI-H-BI1-afe-g-08-0-UI	AW206294	514
-	gb_gss1:AQ585600	1.2e+03	115.80	7.00	RPCI-11-435I1.TV RPCI-	AQ585600	514
-	gb_est30:AI663457	1.2e+03	115.79	7.00	uk33a11.y1 Sugano mous	AI663457	515
-	gb_est22:AI069734	1.2e+03	115.76	7.00	TENU2778 T. cruzi epim	AI069734	517
-	gb_est36:AI929639	1.2e+03	115.76	7.00	au61h05.y1 Schneider_L	AI929639	517
-	gb_est36:AI912708	1.2e+03	115.75	7.00	wei9i12.x1 NCI_CGAP_Lu	AI912708	518
-	gb_est40:AA861815	1.2e+03	115.73	7.00	ak35g07.s1 Soares_test	AA861815	519
-	gb_est46:AW370475	1.2e+03	115.73	7.00	RC2-BT0256-111199-012-	AW370475	519
-	gb_est33:AI772785	1.2e+03	115.72	7.00	EST253885 tomato resis	AI772785	520
-	gb_est36:AI931250	1.2e+03	115.72	7.00	u158g08.y1 Sugano mous	AI931250	520
-	gb_est36:AI931644	1.2e+03	115.71	7.00	u117g01.y1 Sugano mous	AI931644	520
-	gb_est36:AI931208	1.2e+03	115.71	7.00	u157g03.y1 Sugano mous	AI931208	521
-	gb_est44:AQ703085	1.2e+03	115.71	7.00	HS_5466_A2_F10_T7A RPC	AQ703085	521
-	gb_gss4:AQ783295	1.2e+03	115.71	7.00	HS_5445_A2_E10_T7A RPC	AQ783295	521
-	gb_est34:AI806813	1.2e+03	115.69	7.00	wf15f03.x1 Soares_NFL_	AI806813	522
-	gb_est46:AW318748	1.2e+03	115.69	7.00	un05g05.y1 Sugano mous	AW318748	522
-	gb_gss11:AQ189315	1.2e+03	115.69	7.00	HS_3199_B1_A11_T7 CIT	AQ189315	522
-	gb_est33:AQ533680	1.2e+03	115.68	7.00	RPCI-11-384K11.TV RPCI	AQ533680	523
-	gb_est18:T21750	1.2e+03	115.67	7.00	3758 Lambda-PRL2 Arabidc	T21750	524
-	gb_gss11:AQ175573	1.2e+03	115.67	7.00	CITBI-E1-2553M20.TR CI	AQ175573	524
-	gb_gss16:AQ602724	1.2e+03	115.65	7.00	HS_2116_A2_G09_T7C CII	AQ602724	525
-	gb_est15:AI540285	1.2e+03	115.64	7.00	tg34a09.x1 NCI_CGAP_Ut	AI540285	526
-	gb_est10:AI400162	1.2e+03	115.63	7.00	tg67g09.x1 Soares_NhHM	AI400162	527
-	gb_est38:AW092660	1.2e+03	115.61	7.00	EST285840 tomato mixed	AW092660	528
-	gb_est20:AA869046	1.2e+03	115.60	7.00	vq30b01.r1 Barstead_st	AA869046	529
+	gb_gss11:AQ394579	1.2e+03	115.60	7.00	CITBI-E1-2553M20.TR CI	AQ394579	529
-	gb_est11:AA266644	1.2e+03	115.58	7.00	mr2ld08.r1 Soares mous	AA266644	530
-	gb_est30:AI648233	1.2e+03	115.57	7.00	uk40d12.x1 Sugano mous	AI648233	530
-	gb_est38:AW128873	1.2e+03	115.54	7.00	xe89c02.x1 NCI_CGAP_Br	AW128873	531
-	gb_est38:AW091882	1.2e+03	115.53	7.00	EST285062 tomato mixed	AW091882	533
-	gb_est32:AI722315	1.2e+03	115.52	7.00	fc32e03.x1 Zebrafish W	AI722315	534
+	gb_gss13:AQ336107	1.2e+03	115.50	7.00	HS_5019_B1_C10_T7 RPCI	AQ336107	536

Right block

flag	id	val	score		description	accession	#
.	gb_est21:AA986571	1.2e+03	115.46	7.00	uc81e06.y1 Sugano m	AA986571	539
.	gb_est9:AA116194	1.2e+03	115.45	7.00	rs01f12.r1 Sommer P	AA116194	540
.	gb_est22:AI049371	1.2e+03	115.45	7.00	uc83h12.y1 Sugano m	AI049371	540
.	gb_est24:AI570810	1.2e+03	115.45	7.00	tr67012.x1 NCI_CGAP	AI570810	540
.	gb_gss4:AQ697192	1.2e+03	115.44	7.00	HS_5519_B2_A04_SP6E	AQ697192	541
.	gb_est20:AA864874	1.2e+03	115.40	7.00	oh03d08.s1 NCI_CGAP	AA864874	544
.	gb_est2:R11894	1.2e+03	115.39	7.00	yf50a09.r1 Soares inf	R11894	545
.	gb_gss3:B65816	1.2e+03	115.37	7.00	CIT-HSP-2023C22.TF CI	B65816	546
.	gb_est39:AW142037	1.2e+03	115.37	7.00	EST292243 Normalize	AW142037	546
.	gb_gss11:AQ281315	1.2e+03	115.36	7.00	RPCI11-91E15.TV RPC	AQ281315	547
.	gb_est21:C9180	1.2e+03	115.36	7.00	C9180 Dictyostelium	C9180	547
.	gb_gss1:FR0012887	1.2e+03	115.35	7.00	F.rubripes GSS sequ	AL004137	547
.	gb_gss9:AQ015726	1.2e+03	115.35	7.00	RPCI-23-282F9.TY.RP	AQ015726	548
.	gb_gss5:AQ827592	1.3e+03	115.33	7.00	HS_5472_A2_F12_T7A	AQ827592	549
.	gb_gss4:AQ701932	1.3e+03	115.32	7.00	HS_5396_B2_E02_T7A	AQ701932	550
.	gb_gss13:AQ430015	1.3e+03	115.31	7.00	HS_5061_B2_D03_T7A	AQ430015	551
.	gb_est17:AA617054	1.3e+03	115.31	7.00	v121c04.r1 Barstead	AA617054	551
.	gb_est23:AI088606	1.3e+03	115.31	7.00	qb14e01.x1 Soares_p	AI088606	552
.	gb_est29:AI976951	1.3e+03	115.29	7.00	zp09h11.r1 Stratage	AI976951	552
.	gb_est10:AA167479	1.3e+03	115.29	7.00	HS_3224_A2_E12_T7 C	AA167479	552
.	gb_gss11:AQ189784	1.3e+03	115.28	7.00	HS_5071_A1_H12_T7 R	AQ189784	553
.	gb_est25:AI279395	1.3e+03	115.27	7.00	qx66b07.x1 NCI_CGAP	AI279395	554
.	gb_gss13:AQ357745	1.3e+03	115.26	7.00	CITBI-E1-2537H13.TF	AQ357745	555
.	gb_gss13:AI734529	1.3e+03	115.26	7.00	606032F08.x1 606 ·	AI734529	555
.	gb_gss3:B68832	1.3e+03	115.23	7.00	CIT-HSP-2024B19.TR CI	B68832	557
.	gb_est22:AI006180	1.3e+03	115.23	7.00	ua87g03.r1 Soares_m	AI006180	557
.	gb_est38:AI939456	1.3e+03	115.22	7.00	tf19f01.x5 NCI_CGAP	AI939456	558
.	gb_est27:AW049416	1.3e+03	115.20	7.00	UI-M-BH1-ane-g-12-0	AW049416	559
.	gb_est27:AI414985	1.3e+03	115.20	7.00	mb46c10.x1 Soares m	AI414985	559
.	gb_est31:AI586070	1.3e+03	115.20	7.00	vu91cll.x1 Stratage	AI586070	559
.	gb_est21:AI672704	1.3e+03	115.18	7.00	we57f06.x1 Soares_t	AI672704	561
.	gb_est32:AA986908	1.3e+03	115.17	7.00	uc80a03.y1 Sugano m	AA986908	562
.	gb_est4:AI745947	1.3e+03	115.17	7.00	605077G12.x1 605 ·	AI745947	562
.	gb_gss7:AW173310	1.3e+03	115.17	7.00	xj85h03.x1 Soares_N	AW173310	562
.	gb_est23:AQ950822	1.3e+03	115.15	7.00	Sheared DNA-52C20.T	AQ950822	563
.	gb_est33:AA470014	1.3e+03	115.15	7.00	zu10f04.r1 Soares_t	AA470014	563
.	gb_gss7:AI097495	1.3e+03	115.14	7.00	qb90dd04.x1 Soares_f	AI097495	564
.	gb_gss11:AI774683	1.3e+03	115.14	7.00	EST255783 tomato re	AI774683	564
.	gb_gss15:AQ925711	1.3e+03	115.12	7.00	RPCI-23-297O11.TJ R	AQ925711	566
.	gb_est9:AQ156765	1.3e+03	115.12	7.00	nbxb0008K18f CUGI R	AQ156765	566
.	gb_est17:AQ591466	1.3e+03	115.09	7.00	HS_5413_A1_H02_T7A	AQ591466	569
.	gb_est37:AA541729	1.3e+03	115.08	7.00	n19d02.s1 NCI_CGAP	AA541729	569
.	gb_est30:AA068964	1.3e+03	115.08	7.00	mm61f10.r1 Stratage	AA068964	571
.	gb_gss8:C77164	1.3e+03	115.05	7.00	Mouse 3.5-dpc	C77164	571
.	gb_gss13:AA726464	1.3e+03	115.05	7.00	Vu40e08.r1 Barstead	AA726464	573
.	gb_est39:AI977956	1.3e+03	115.03	7.00	49603G01.x1 496 ·	AI977956	573
.	gb_est25:AL048988	1.3e+03	115.02	7.00	DKF2p434N1018_s1 43	AL048988	574
.	gb_gss8:AQ783295	1.3e+03	114.99	7.00	RPCI23-N1018.TJ RP	AQ783295	576
.	gb_gss13:AQ398831	1.3e+03	114.98	7.00	mgxb0015C01f CUGI R	AQ398831	577
.	gb_est39:AW106921	1.3e+03	114.98	7.00	um34c05.x1 Sugano mou	AW106921	577
.	gb_est25:AI309768	1.3e+03	114.98	7.00	qo75b03.x1 NCI_CGAP	AI309768	577
.	gb_est9:AI726669	1.3e+03	114.97	7.00	BNLGH5789 Six-day	AI726669	578
.	gb_gss13:AQ368594	1.3e+03	114.95	7.00	HS_5037_A2_G07_SP6E	AQ368594	579
.	gb_gss13:AQ390093	1.3e+03	114.95	7.00	RPCI11-146E23.TV RP	AQ390093	579
.	gb_gss15:AW068627	1.3e+03	114.94	7.00	cn21e12.x1 Normal H	AW068627	580
.	gb_gss13:AW246217	1.3e+03	114.92	7.00	2821619.5prime NIH	AW246217	582
.	gb_est19:AI166936	1.3e+03	114.91	7.00	xylem.est.719 Popla	AI166936	582
.	gb_est19:AI259708	1.3e+03	114.91	7.00	nbxb0023L09r CUGI R	AI259708	582
.	gb_est9:AA518350	1.3e+03	114.89	7.00	RPCI11-134O21.TJ RP	AA518350	583
.	gb_gss12:AA628539	1.3e+03	114.88	7.00	v110b12.r1 Barstead	AA628539	583
.	gb_gss13:AI956374	1.3e+03	114.88	7.00	af27h06.s1 Soares_t	AI956374	584
.	gb_est7:W65917	1.3e+03	114.88	7.00	me09d09.r1 Soares mou	W65917	585
.	gb_est24:AQ273510	1.3e+03	114.87	7.00	EST225618 Normalize	AQ273510	585
.	gb_est11:AI228923	1.3e+03	114.87	7.00	HS_2014_A2_H01_MR C	AI228923	586
.	gb_est38:AQ901624	1.3e+03	114.87	7.00	cn21a03.r1 Normal H	AQ901624	586
.	gb_gss15:AW068530	1.3e+03	114.86	7.00	RPCI-11-422M11.TV R	AW068530	586
.	gb_est10:AQ554295	1.3e+03	114.83	7.00	vo98a12.r1 Soares_m	AQ554295	587
.	gb_gss11:AA764479	1.3e+03	114.81	7.00	mp71e08.r1 Soares_t	AA764479	587
.	gb_gss12:AA125356	1.3e+03	114.81	7.00	vs88e06.r1 Barstead	AA125356	589
.	gb_est19:AA792586	1.3e+03		7.00	nbxb0043D01r CUGI R	AA792586	591
.	gb_gss12:AQ328604					AQ328604	591

Tue Jun 13 10:29:59 2000

us-09-252-691-7056.oligo.rst

Left block

±	ID	Description	N	E-value	Score	Wt
-	gb_est23:AI14963	AI114963 u141all.y1 Sugano mous	828	1.8e+03	112.40	7.00
-	gb_gss3:B19885	B19885 T18N7-Sp6 TAMU Arabidops	833	1.8e+03	112.36	7.00
+	gb_est34:AI789173	AI789173 uk51h07.y1 Sugano mous	834	1.8e+03	112.35	7.00
-	gb_gss1:AQ780692	AQ780692 HS_3169_A1_C01_T7C CIT	837	1.8e+03	112.33	7.00
+	gb_gss5:AQ749772	AL053482 Drosophila melanogaste	846	1.9e+03	112.25	7.00
+	gb_gss5:AQ853523	AQ749772 HS_5573_A2_E05_T7A RPC	847	1.9e+03	112.24	7.00
+	gb_est28:AI527753	AQ853523 nbxb0021C20f CUGI Rice	853	1.9e+03	112.19	7.00
-	gb_gss15:AQ576187	AI527753 uj28g03.y1 Sugano mous	859	1.9e+03	112.14	7.00
+	gb_gss5:AQ749934	AQ576187 nbxb0088P23f CUGI Rice	873	1.9e+03	112.03	7.00
+	gb_est38:AW054104	AQ749934 HS_5576_A1_C04_SP6 RPC	876	1.9e+03	112.00	7.00
+	gb_est31:AU067178	AW054104 L30-2085T3 Ice plant I	881	1.9e+03	111.96	7.00
+	gb_gss1:CNS0051B	AU067178 AU067178 Sugano mouse	882	1.9e+03	111.95	7.00
-	gb_est25:AI323216	AL059536 Drosophila melanogaste	888	1.9e+03	111.90	7.00
+	gb_gss4:AQ688009	AI323216 mp71e08.y1 Soares_thym	913	2.0e+03	111.71	7.00
+	gb_gss1:B11110	AQ688009 nbxb0076F08r CUGI Rice	919	2.0e+03	111.66	7.00
+	gb_gss6:AQ895470	B11110 T7G13-T7 TAMU Arabidops	920	2.0e+03	111.65	7.00
-	gb_gss2:CNS0188V	AQ895470 HS_3117_A2_A01_MR CIT	954	2.1e+03	111.39	7.00
+	gb_est7:W40957	AL109033 Drosophila melanogaste	966	2.1e+03	111.30	7.00
-	gb_gss2:CNS017C9	W40957 mc34e11.r1 Soares mouse	979	2.1e+03	111.21	7.00
-	gb_gss1:CNS005MA	AL107859 Drosophila melanogaste	989	2.1e+03	111.14	7.00
-	gb_gss1:CNS000FP	AL067244 Drosophila melanogaste	1027	2.2e+03	110.87	7.00
-	gb_gss1:CNS000NB	AL052862 Drosophila melanogaste	1049	2.3e+03	110.72	7.00
-	gb_gss1:CNS001IX	AL063259 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss1:CNS0024A	AL061064 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss1:CNS002RZ	AL061833 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss1:CNS0033P	AL063292 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss1:CNS00E3V	AL063815 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss1:CNS00K4Z	AL068708 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss3:B10533	AL077231 Drosophila melanogaste	1101	2.4e+03	110.37	7.00
-	gb_gss3:B11549	B10533 F12-Sp6 IGF Arabidops	1163	2.5e+03	109.98	7.00
+	gb_est24:AI221747	B11549 F26P9-Sp6 IGF Arabidop	1208	2.5e+03	109.71	7.00
+	gb_est9:AA124006	AI221747 qg93f02.x1 Soares_NFL_T	34	1.2e+03	115.82	6.00
+	gb_est19:AA757476	AA124006 mq10a06.r1 Barstead MPI	49	1.6e+03	113.22	6.00
+	gb_est46:AW318967	AA757476 zg41c04.s1 Soares_pinea	49	1.6e+03	113.22	6.00
+	gb_est25:AW265471	AW318967 un10b04.y1 Sugano mouse	50	1.7e+03	113.07	6.00
+	gb_est18:AA730665	AW265471 xq69f03.x1 NCI_CGAP_HN7	57	1.9e+03	112.14	6.00
+	gb_est18:AA240825	AA730665 nw46g03.s1 NCI_CGAP_Ewi	58	1.9e+03	112.02	6.00
+	gb_est21:AA990602	AA240825 mw17e06.r1 Soares mouse	63	2.1e+03	111.43	6.00
+	gb_est20:AA905978	AA990602 ua65g06.s1 Soares_thymu	66	2.2e+03	111.09	6.00
+	gb_est20:AA822882	AA905978 oj89c03.s1 Soares_NFL_I	67	2.2e+03	110.99	6.00
+	gb_est18:AA716326	AA822882 vq91g12.r1 Knowles Solt	70	2.3e+03	110.67	6.00
+	gb_est14:AA413641	AA716326 zg19g02.s1 Soares_fetal	75	2.4e+03	110.18	6.00
+	gb_est4:H40990	AA413641 vc57b02.s1 Knowles Solt	77	2.4e+03	110.00	6.00
+	gb_est30:AA030267	H40990 yn96d02.s1 Soares adult b	78	2.5e+03	109.81	6.00
+	gb_est30:AI663950	AA030267 m125b06.r1 Soares mouse	79	2.5e+03	109.81	6.00
+	gb_est44:AW191390	AI663950 uj46h04.x1 Sugano mouse	79	2.5e+03	109.72	6.00
+	gb_est37:AI969589	AW191390 suc204 Eucalypt cambium	80	2.6e+03	109.46	6.00
+	gb_est4:AQ060374	AI969589 wz69a07.x1 NCI_CGAP_Me	83	2.7e+03	109.21	6.00
+	gb_est4:H55369	AQ060374 CIT-HSP-2349E3.TF CIT-	86	2.9e+03	108.80	6.00
+	gb_est14:AA433749	H55369 CHR220308 Chromosome 22	91	2.9e+03	108.80	6.00
+	gb_est15:AA529216	AA433749 v56b01.r1 Soares mouse	91	2.9e+03	108.65	6.00
+	gb_est27:AI407374	AI313603 EST057 Mouse ES cell La	93	3.0e+03	108.65	6.00
+	gb_est25:AI267774	AA529216 v133f10.r1 Soares mouse	93	3.0e+03	108.57	6.00
+	gb_est22:AU010782	AI407374 EST235663 Normalized ra	94	3.0e+03	108.42	6.00
+	gb_est46:AW354923	AA207832 mv79h04.r1 GuayWoodford	96	3.0e+03	108.35	6.00
-	gb_est7:W64053	AI910351 ok83f04.s1 NCI_CGAP_Kid	97	3.1e+03	108.13	6.00
+	gb_est28:AI471311	AI416225 fb20b04.y1 Zebrafish Wa	100	3.2e+03	108.13	6.00
+	gb_est45:AW259522	AA092430 119427.seq.F Human fet	100	3.2e+03	108.13	6.00
+	gb_est20:AA873029	AW153181 se37a02.y1 Gm-c1015 GI	101	3.2e+03	108.06	6.00
		AW159131 za62a04.s3 Xenopus EST	103	3.3e+03	107.92	6.00
		AI663950 uj46h04.x1 Sugano mouse	104	3.3e+03	107.85	6.00
		AQ060374 CIT-HSP-2349E3.TF CIT-	104	3.3e+03	107.85	6.00
		AA433749 v56b01.r1 Soares mouse	105	3.3e+03	107.78	6.00
		AA529216 v133f10.r1 Soares mouse	105	3.3e+03	107.78	6.00
		AI407374 EST235663 Normalized ra	106	3.3e+03	107.72	6.00
		AI267774 ap64h11.x1 Stanley Frc	106	3.4e+03	107.72	6.00
		AU010782 AU010782 Schizosacchar	106	3.4e+03	107.52	6.00
		AW354923 37609 MARC 1BOV Bos ta	109	3.4e+03	107.52	6.00
		W64053 md77d11.r1 Soares mouse	109	3.4e+03	107.52	6.00
		AI471311 ta14h03.x1 NCI_CGAP_Ly	109	3.5e+03	107.39	6.00

Right block

ID	Description	N	E-value	Score	Wt	±	Ref
AQ341247	RPCI11-119N21.TV RP	111	3.5e+03	107.39	6.00	+	gb_gss12:AQ341247
AA014128	mh28h01.x1 Soares m	111	3.5e+03	107.32	6.00	-	gb_est8:AA014128
AA484507	nf08f01.s1 NCI_CGAP	112	3.5e+03	107.32	6.00	+	gb_est15:AA484507
AA711354	vt77b08.r1 Barstead	112	3.5e+03	107.32	6.00	-	gb_est26:AA711354
AI366474	ao83d11.x1 Stanley	112	3.5e+03	107.32	6.00	+	gb_est26:AI366474
AW259612	up31a12.x1 NCI_CGAP	112	3.5e+03	107.26	6.00	+	gb_est45:AW259612
AA484423	nf07b04.s1 NCI_CGAP	113	3.6e+03	107.26	6.00	+	gb_est15:AA484423
AA484509	nf08f03.s1 NCI_CGAP	113	3.6e+03	107.20	6.00	+	gb_est15:AA484509
N86349	J7848F Human fetal he	114	3.6e+03	107.13	6.00	-	gb_est6:N86349
AQ042290	CIT-HSP-2330G3.TF C	115	3.6e+03	107.13	6.00	-	gb_gss3:AQ042290
D18117	MUSGS00362 Mouse live	116	3.6e+03	107.07	6.00	-	gb_est5:D18117
AA757093	ah55e10.s1 Soares_t	116	3.6e+03	107.07	6.00	+	gb_est19:AA757093
AI015107	ot90b12.s1 Soares_t	116	3.6e+03	107.07	6.00	-	gb_est21:AI015107
H92931	yt93g10.s1 Soares_pin	117	3.6e+03	107.01	6.00	+	gb_est5:H92931
B22317	F18J6rF IGF Arabidops	117	3.7e+03	107.01	6.00	-	gb_gss3:B22317
AA484425	nf07b06.s1 NCI_CGAP	118	3.7e+03	106.95	6.00	+	gb_est16:AA484425
AA549126	vk74d04.s1 Knowles	118	3.7e+03	106.95	6.00	-	gb_est18:AA549126
AA736673	oa22e06.s1 NCI_CGAP	118	3.7e+03	106.95	6.00	+	gb_est18:AA736673
D45337	HUMHGI5111 Human cere	119	3.7e+03	106.89	6.00	-	gb_est5:D45337
AA501682	ng0e03.s1 NCI_CGAP	119	3.7e+03	106.89	6.00	+	gb_est15:AA501682
AA866045	oh52g08.s1 NCI_CGAP	119	3.7e+03	106.89	6.00	+	gb_est20:AA866045
AQ585865	RPCI-11-432E1.TJ RP	119	3.7e+03	106.83	6.00	-	gb_gss9:AQ585865
AA596402	vo38d09.r1 Barstead	120	3.7e+03	106.83	6.00	-	gb_est16:AA596402
AW347413	3I109 MARC 2PIG Sus	120	3.7e+03	106.83	6.00	+	gb_est46:AW347413
AQ974158	CIT-23-331H10.TJ R	120	3.7e+03	106.83	6.00	-	gb_gss7:AQ974158
AQ056587	CIT-HSP-2341D8.TF C	120	3.7e+03	106.83	6.00	-	gb_gss9:AQ056587
AQ077099	CIR-HSP-2359F11.TR	120	3.7e+03	106.77	6.00	-	gb_gss9:AQ077099
R53268	yg82a03.r1 Soares inf	121	3.8e+03	106.77	6.00	-	gb_est5:R53268
AA642376	ns29e03.s1 NCI_CGAP	121	3.8e+03	106.77	6.00	+	gb_est17:AA642376
AQ930968	RPCI-23-255F5.TV RP	121	3.8e+03	106.71	6.00	-	gb_gss7:AQ930968
AA484584	nf10b02.s1 NCI_CGAP	121	3.8e+03	106.71	6.00	+	gb_est15:AA484584
AA503773	ng07h06.s1 NCI_CGAP	122	3.8e+03	106.71	6.00	+	gb_est15:AA503773
B35883	HS-1031-A2-B11-MR.ab1	122	3.8e+03	106.66	6.00	-	gb_gss3:B35883
AI950826	wx55f09.x1 NCI_CGAP	122	3.8e+03	106.66	6.00	+	gb_est37:AI950826
B82206	RPCI11-13C4.TVB RPCI-	123	3.8e+03	106.60	6.00	-	gb_gss3:B82206
AA611735	v089g09.r1 Barstead	124	3.8e+03	106.60	6.00	-	gb_est17:AA611735
AA620681	af89b06.s1 Soares_t	124	3.8e+03	106.60	6.00	+	gb_est17:AA620681
AA852135	NHTBCae10e02f1 Norm	124	3.8e+03	106.60	6.00	+	gb_est20:AA852135
AA902765	ok72a01.s1 NCI_CGAP	124	3.9e+03	106.54	6.00	+	gb_est25:AA902765
AI321028	d2h07nm.r1 Neurospo	124	3.9e+03	106.54	6.00	-	gb_est31:AI321028
AA036515	m159g11.r1 Soares m	125	3.9e+03	106.48	6.00	+	gb_est8:AA036515
AA545426	vj95h02.s1 Knowles	125	3.9e+03	106.48	6.00	+	gb_est15:AA545426
AI919204	tu32d06.x1 NCI_CGAP	126	3.9e+03	106.48	6.00	+	gb_est36:AI919204
AV392280	AV392280 Chlamydomo	126	3.9e+03	106.43	6.00	-	gb_est44:AV392280
AQ372710	RPCI11-157E2.TV RPC	126	3.9e+03	106.43	6.00	-	gb_gss13:AQ372710
AA152650	mq55c07.r1 Soares_t	127	4.0e+03	106.43	6.00	+	gb_est15:AA152650
AA484382	nf06f11.s1 NCI_CGAP	127	4.0e+03	106.37	6.00	+	gb_est15:AA484382
AQ073803	EP(3)3186-3prime Dr	127	4.0e+03	106.37	6.00	-	gb_gss9:AQ073803
N61751	TgESTzy29f02.r1 TgRH	127	4.0e+03	106.37	6.00	-	gb_est6:N61751
AA167932	ms21d08.r1 Stratage	128	4.0e+03	106.37	6.00	+	gb_est15:AA167932
AA484594	nf10c06.s1 NCI_CGAP	128	4.0e+03	106.37	6.00	+	gb_est15:AA484594
AA557247	nl75e08.s1 NCI_CGAP	128	4.0e+03	106.37	6.00	+	gb_est21:AA557247
AA973289	on94b04.s1 Soares_N	128	4.0e+03	106.26	6.00	+	gb_est31:AA973289
AI708130	as99h10.x1 Barstead	128	4.0e+03	106.26	6.00	+	gb_est11:AI708130
AI506278	vq55f01.x1 Barstead	129	4.0e+03	106.21	6.00	-	gb_est6:AI506278
AA249231	hfe019.seq.F Human	130	4.0e+03	106.21	6.00	+	gb_est11:AA249231
N87868	K0440F Human fetal he	131	4.0e+03	106.15	6.00	-	gb_est6:N87868
N62010	TgESTzy20f06.r1 TgRH	132	4.1e+03	106.15	6.00	-	gb_est15:N62010
AA469337	nc70e07.s1 NCI_CGAP	132	4.1e+03	106.15	6.00	+	gb_est15:AA469337
AW150339	wa18f10.x1 NCI_CGAP	132	4.1e+03	106.15	6.00	+	gb_est30:AW150339
AV029631	AV029631 Mus muscul	132	4.1e+03	106.15	6.00	-	gb_est35:AV029631
AI822421	L0-88T3 Ice plant	132	4.1e+03	106.10	6.00	+	gb_est15:AI822421
AA054554	zk83g04.s1 Soares_p	133	4.1e+03	106.10	6.00	+	gb_est8:AA054554
AA833637	aj45e03.s1 Soares_t	133	4.1e+03	106.10	6.00	+	gb_est15:AA833637
AI097861	ue40g07.y1 Sugano m	133	4.1e+03	106.10	6.00	+	gb_est23:AI097861
AW151370	xe75a07.x1 NCI_CGAP	133	4.1e+03	106.10	6.00	+	gb_est40:AW151370
T25525	EST00558 Unequalized	134	4.1e+03	106.04	6.00	-	gb_est11:T25525
AA249363	j1473.seq.F Human f	134	4.1e+03	106.04	6.00	+	gb_est11:AA249363
AV073190	AV073190 Mus muscul	135	4.1e+03	105.99	6.00	-	gb_est33:AV073190
AA565750	o123h12.s1 NCI_CGAP	136	4.2e+03	105.94	6.00	+	gb_est17:AA565750
AA611292	v17h07.r1 Knowles	136	4.2e+03	105.94	6.00	-	gb_est17:AA611292
B28459	F24L12TR IGF Arabidop	136	4.2e+03	105.94	6.00	+	gb_gss3:B28459
AA420637	nc6if10.r1 NCI_CGAP	137	4.2e+03	105.89	6.00	+	gb_est14:AA420637

```
gb_est18:AA674007    -    6.00  105.89  4.2e+03  !  AA674007 vp70a10.r1 Knowles So:
gb_est18:AA687136    -    6.00  105.89  4.2e+03  !  AA687136 nv63e02.s1 NCI_CGAP_GC
gb_est38:AW082321    +    6.00  105.89  4.2e+03  !  AW082321 xb65d10.x1 Soares_NFL_
gb_gss9:AQ076792          6.00  105.89  4.2e+03  !  AQ076792 CIT-HSP-236I15.TR CIT-
gb_est15:AA480572    -    6.00  105.83  4.2e+03  !  AA480572 nc70e07.r1 NCI_CGAP_Pi
gb_est19:AA775657    +    6.00  105.78  4.3e+03  !  AA775657 zf26d10.s1 Soares_feta
gb_est45:AW243485    +    6.00  105.78  4.3e+03  !  AW243485 xm95h06.x1 NCI_CGAP_Ki
gb_est14:AA409493    +    6.00  105.73  4.3e+03  !  AA409493 EST01255 Mouse 7.5 dpc
gb_est16:AA597233    +    6.00  105.73  4.3e+03  !  AA597233 vo28c04.r1 Barstead mc
gb_gss1:FR022013     +    6.00  105.73  4.3e+03  !  AL014884 F.rubripes GSS sequenc
gb_est21:AA948266    +    6.00  105.68  4.3e+03  !  AA948266 oq34a05.s1 NCI_CGAP_GC
gb_est15:AA483254    +    6.00  105.63  4.3e+03  !  AA483254 nf03f10.s1 NCI_CGAP_Li
```

seq_name: gb_est26:AI373988

```
seq_documentation_block:
LOCUS       AI373988    283 bp    mRNA    EST    15-JAN-1999
DEFINITION  SWOvAFCAP32B02SK Onchocerca volvulus adult female cDNA
            (SAW98MLW-OvAF) Onchocerca volvulus cDNA clone SWOvAFCAP32B02 5',
            mRNA sequence.
ACCESSION   AI373988
VERSION     AI373988.1  GI:4160024
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
ORGANISM    Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 283)
AUTHORS     Lizotte-Waniewski,M. and Williams,S.A.
TITLE       Genes expressed in adult female stage of Onchocerca volvulus
JOURNAL     Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797276.
            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES            Location/Qualifiers
     source         1..283
                    /organism="Onchocerca volvulus"
                    /db_xref="taxon:6282"
                    /clone="SWOvAFCAP32B02"
                    /clone_lib="Onchocerca volvulus adult female cDNA
                    (SAW98MLW-OvAF)"
                    /sex="female"
                    /dev_stage="adult"
                    /lab_host="Xl1-Blue MRF'"
                    /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                    Xho I; Filarial nematode parasite of humans. Two adult
                    female worms of Onchocerca volvulus were isolated from
                    consenting patients and quick frozen. Adult female mRNA
                    was converted to double-stranded cDNA using reverse
                    transcriptase and oligo(dT) followed by RNase H and DNA
                    pol I. The library has 7 x 10E5 independent recombinants
                    and the average  insert size is ~1100bp.The library was
                    constructed by Michelle Lizotte-Waniewski with worms
                    provided by Dr. Sara Lustigman. The library is available
                    from Dr.Steven A. Williams, email: genome@smith.edu."
BASE COUNT      63 a      86 c      74 g      60 t
```

```
alignment_scores:
Quality:      52.00         Length:      52
Ratio:         1.000        Gaps:         0
Percent Similarity: 100.000  Percent Identity: 100.000
```

```
alignment_block:
US-09-252-691-7056 x AI373988
```

Algn seg 1/1 to: AI373988 from: 1 to: 283

```
151 TrpLeuTrpProArgAsnProProIleArgGluArgLysSerIleProTh 167
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 87 TGGTTATGGCCGCCGGAATCCACCAATTCGTGAACGCAAAAGTATTCCCAC 136

167 rSerTrpLeuLysIleThrLeuTyrGluGlyArgAsnArgGlnValArgA 184
    |||||||||||||||||||||||||||||||||||||||||||||||||||
137 CAGCTGGCTGAAGATCACCTTATATGAAGGACGTAATCGCCAGGTGCGCC 186

184 rgMetThrArgHisValGlyPheProThrLeuArgLeuIleArgTyrAla 200
    |||||||||||||||||||||||||||||||||||||||||||||||||||
187 GCATGACGCGCCCATGTTGGCTTCCCCACGCTGCGACTGATTCGCTATGCG 236

201 MetGly 202
    ||||||
237 ATGGGT 242
```

seq_name: em_est20:AW017830

```
seq_documentation_block:
ID   AW017830   standard; RNA; EST; 295 BP.
XX
AC   AW017830;
XX
SV   AW017830.1
XX
DT   14-SEP-1999 (Rel. 61, Created)
DT   14-SEP-1999 (Rel. 61, Last updated, Ver
XX
DE   614064A09.x1 614 - root cDNA library f
DE   mRNA sequence.
XX
KW   EST.
XX
OS   Zea mays
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Zea.
XX
RN   [1]
RP   1-295
RA   Walbot V.;
RT   "Maize ESTs from various cDNA libraries sequenced at Stanford University";
RL   Unpublished.
XX
DR   MaizeDB; Probe/236176; p-std61406409.
XX
CC   Contact: Walbot V
CC   Department of Biological Sciences
CC   Stanford University
CC   855 California Ave, Palo Alto, CA 94304, USA
CC   Tel: 650 723 2227
CC   Fax: 650 725 8221
CC   Email: walbot@stanford.edu
CC   Plate: 614064 row: A  column: 09.
XX
FH   Key            Location/Qualifiers
FH
FT   source         1..295
FT                  /db_xref="taxon:4577"
FT                  /db_xref="ESTLIB:1834"
FT                  /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
FT                  EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
FT                  Lab (LM)"
FT                  /organism="Zea mays"
FT                  /cultivar="W23"
FT                  /clone_lib="614 - root cDNA library from Walbot Lab"
FT                  /tissue_type="root"
FT                  /dev_stage="3-4 days old"
FT                  /lab_host="XlOLR"
XX
```

NA,

SQ   Sequence 295 BP; 80 A; 76 C; 76 G; 63 T; 0 other;

alignment_scores:
        Quality: 25.00   Length: 25
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AW017830

Algn seg 1/1 to: AW017830 from: 1 to: 295

    68 PheIleProValGlnGlyValTyrAlaAlaGlyArgLeuAspArgAspSe 84
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   146 TTCATCCCGGTTCAGGGTGTTTATGCAGCAGGTCGCCTTGACCCGCGATAG 195

    84 rGluGlyLeuLeuValLeuThrAsn 92
       |||||||||||||||||||||||||
   196 CGAAGGGTTGCTCGTGGTGCTGACCAAT 220

seq_name: gb_gss11:AQ212974

seq_documentation_block:
LOCUS   AQ212974   546 bp   DNA   GSS   18-SEP-1998
DEFINITION   HS_3214_B2_E03_MR CIT Approved Human Genomic Sperm Library D Homo
             sapiens genomic clone Plate=3214 Col=6 Row=J, genomic survey
             sequence.
ACCESSION   AQ212974
VERSION   AQ212974.1 GI:3624175
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 546)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and
             scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   On Sep 10, 1998 this sequence version replaced gi:3554741.
             Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Sequence Tagged Connector
             Plate: 3214 Row: J column: 6
             Class: BAC ends
             High quality sequence stop: 546.
FEATURES   Location/Qualifiers
     source 1..546
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="Plate=3214 Col=6 Row=J"
             /clone_lib="CIT Approved Human Genomic Sperm Library D"
             /sex="male"
             /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
             E-Coli DH10B"
BASE COUNT   144 a   107 c   120 g   168 t   7 others
ORIGIN

alignment_scores:
        Quality: 10.00   Length: 10
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

---

alignment_block:
US-09-252-691-7056 x AQ212974/rev ..

Algn seg 1/1 to reverse of: AQ212974 from: 1 to: 546

    94 GlyValLeuGlnAlaArgLeuThrGlnPro 103
       ||||||||||||||||||||||||||||||
   408 GGTGTCTGCAGGCAAGGCTGACACAGCCN 379

seq_name: gb_est11:AA237292

seq_documentation_block:
LOCUS   AA237292   160 bp   mRNA   EST   03-MAR-1997
DEFINITION   mw96d12.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:678551 5'
             similar to TR:G478984 G478984 PEROXISOMAL ENOYL HYDRATASE-LIKE
             PROTEIN. ;, mRNA sequence.
ACCESSION   AA237292
VERSION   AA237292.1 GI:1861313
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 160)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1404819.
             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:418255
             Trace considered overall poor quality
             Possible reversed clone: similarity on wrong strand
             Seq primer: -28m13 rev2 ET from Amersham
             High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source 1..160
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="IMAGE:678551"
             /clone_lib="Soares mouse NML"
             /tissue_type="liver"
             /lab_host="DH10B"
             /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
             polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
             was primed with a Not I - oligo(dT) primer [5'
             TGTTACCAATCTGAAGTGGGAGCCGCGGAATCTTTTTTTTTTTTTTTTT 3'];
             double-stranded cDNA was ligated to Eco RI adaptors
             (Pharmacia), digested with Not I and cloned into the Not I
             and Eco RI sites of the modified pT7T3 vector. Library
             constructed and normalized by Bento Soares and M.Fatima
             Bonaldo."
BASE COUNT   36 a   40 c   41 g   43 t
ORIGIN

alignment_scores:
        Quality: 8.00   Length: 8
        Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:

```
         83 AspSerGluGlyLeuLeuValLeu 90
            ||||||||||||||||||||||||
        207 GACAGTGAAGGTCTTCTGGTACTT 184

seq_name: gb_est39:AV206648

seq_documentation_block:
LOCUS       AV206648       248 bp      mRNA           EST     30-OCT-1999
DEFINITION  AV206648 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 1700088L03 3' similar to X51703 Mouse mRNA for
            ubiquitin, mRNA sequence.
ACCESSION   AV206648
VERSION     AV206648.1   GI:6147501
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 248)
AUTHORS     Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
            Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
            Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
            Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
            Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137194.
            Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
            Hayashizaki,Y.
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES    Location/Qualifiers
    source  1..248
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="1700088L03"
            /clone_lib="RIKEN full-length enriched, adult male testis"
            /sex="male"
            /tissue_type="testis"
            /dev_stage="adult"
            /lab_host="SOLR"
            /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
            prepared and sequenced in Mouse Genome Encyclopedia
            Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in
            RIKEN. Division of Experimental Animal Research in Riken
```

```
US-09-252-691-7056 x AA237292/rev

Align seg 1/1 to reverse of: AA237292 from: 1 to: 160

         83 AspSerGluGlyLeuLeuValLeu 90
            ||||||||||||||||||||||||
         31 GACAGTGAAGGTCTTCTGGTACTT 8

seq_name: gb_est19:AA755409

seq_documentation_block:
LOCUS       AA755409       232 bp      mRNA           EST     21-JAN-1998
DEFINITION  vr53e06.r1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            IMAGE:1124386 3' similar to SW:ECH1_RAT Q62651 PROBABLE PEROXISOMAL
            ENOYL-COA HYDRATASE ;. mRNA sequence.
ACCESSION   AA755409
VERSION     AA755409.1   GI:2802607
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 232)
AUTHORS     Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE       The WashU-HHMI Mouse EST Project
JOURNAL     Unpublished (1996)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900891.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:613722
            High quality sequence stop: 224.
FEATURES    Location/Qualifiers
    source  1..232
            /organism="Mus musculus"
            /strain="B6D2 F1/J"
            /db_xref="taxon:10090"
            /clone="IMAGE:1124386"
            /clone_lib="Knowles Solter mouse 2 cell"
            /tissue_type="embryo"
            /dev_stage="2-cell"
            /lab_host="DH10B"
            /note="Organ: embryo; Vector: pBluescribe (modified);
            Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
            mRNA prepared from 13,500 2-cell stage embryos; Primer:
            SalI(dT); 5'-CGTTCGACCGGTCGACCGTTTTTTTTTTTT-3'; CDNAs
            were cloned into the MluI/SalI sites of a modified
            pBluescribe vector using commercial linkers (NEB)."
BASE COUNT   58 a   61 c   63 g   50 t
ORIGIN

alignment_scores:
    Quality:      8.00        Length:        8
    Ratio:        1.000       Gaps:          0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AA755409/rev

Align seg 1/1 to reverse of: AA755409 from: 1 to: 232
```

contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGACGCGCCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites. "

BASE COUNT   75 a   60 c   41 g   72 t
ORIGIN

alignment_scores:
            Quality: 8.00   Length: 8
             Ratio: 1.000   Gaps: 0
    Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AV206648   ..

Align seg 1/1   to: AV206648   from: 1   to: 248

  29 ArgGlnAlaThrArgAlaThrPro 36
     ||||||||||||||||||||||||
   9 CGCCAAGCAACAAGAGAGAACCCCC 32

seq_name: gb_est12:AA286664

seq_documentation_block:
LOCUS   AA286664   249 bp   mRNA   EST   09-APR-1997
DEFINITION vb79h04.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone
            IMAGE:763255 5' similar to TR:G478984 G478984 PEROXISOMAL ENOYL
            HYDRATASE-LIKE PROTEIN. ;, mRNA sequence.
ACCESSION   AA286664
VERSION   AA286664.1 GI:1931759
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 249)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., and
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:504233.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine₽
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:464175
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 157.
FEATURES   Location/Qualifiers
     source   1..249
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:763255"
                     /clone_lib="Soares mouse 3NME12 5"
                     /sex="unknown"
                     /tissue_type="fetus"
                     /dev_stage="12.5dpc total fetus"

                     /lab_host="DH10B"
                     /note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGAGGCGCCGCCTTATTTTTTTTTTTTTTTTT
                     3'], on total mouse RNA [provided by Minoru Ko, Wayne
                     State Univ.]; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT   57 a   64 c   70 g   58 t
ORIGIN

alignment_scores:
            Quality: 8.00   Length: 8
             Ratio: 1.000   Gaps: 0
    Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AA286664/rev   ..

Align seg 1/1   to reverse of: AA286664   from: 1   to: 249

  83 AspSerGluGlyLeuLeuValLeu 90
     ||||||||||||||||||||||||
  42 GACAGTGAAGGTCTTCTGGTACTT 19

seq_name: gb_est18:AA688759

seq_documentation_block:
LOCUS   AA688759   256 bp   mRNA   EST   12-DEC-1997
DEFINITION vr07b09.r1 Knowles Solter mouse blastocyst B3 Mus musculus cDNA
            clone IMAGE:1111097 5' similar to TR:Q45557 Q45557 ;, mRNA
            sequence.
ACCESSION   AA688759
VERSION   AA688759.1 GI:2678188
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 256)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Nov 6, 1997 this sequence version replaced gi:932290.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:609265
            High quality sequence stop: 165.
FEATURES   Location/Qualifiers
     source   1..256
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1111097"
                     /clone_lib="Knowles Solter mouse blastocyst B3"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"

```
            /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally from mRNA prepared
            from 800 blastocysts.   Primer: SalI(dT):
            5'-CGGTCGACGCCGTGCGACCGTTTTTTTTTTTTTT-3'.   cDNAs were
            cloned into the NotI/SalI sites of a pSPORT vector (Life
            Technologies).  Two different size selections:  B1 (larger
            inserts) and B3."
BASE COUNT      49 a     69 c     77 g     61 t
ORIGIN

alignment_scores:
       Quality:           8.00              Length:          8
       Ratio:             1.000             Gaps:            0
Percent Similarity: 100.000       Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AA688759

Align seg 1/1   to: AA688759  from: 1   to: 256

   180 ArgGlnValArgArgMetThrAla 187
       ||||||||||||||||||||||||
   208 AGACAGGTGCGGAGGATGACAGCT 231

seq_name: gb_est1:X73795

seq_documentation_block:
LOCUS       X73795          271 bp    mRNA         EST       13-NOV-1996
DEFINITION  CHESTM092 Goat mammary gland Capra hircus cDNA, mRNA sequence.
ACCESSION   X73795
VERSION     X73795.1  GI:313488
KEYWORDS    EST.
SOURCE      goat.
ORGANISM    Capra hircus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae;
            Caprinae; Capra.
REFERENCE   1 (bases 1 to 271)
AUTHORS     Le Provost,F., Lepingle,A. and Martin,P.
TITLE       A survey of the goat genome transcribed in the lactating mammary
            gland
JOURNAL     Mamm. Genome 7 (9), 657-666 (1996)
MEDLINE     96359152
COMMENT     Contact: Le Provost.F.
            Laboratoire de Genetique Biochimique
            INRA-CRJ
            78352 Jouy-en-Josas Cedex, FRANCE.
FEATURES    Location/Qualifiers
     source  1..271
            /organism="Capra hircus"
            /db_xref="taxon:9925"
            /clone_lib="Goat mammary gland"
            /note="Mammary gland of Capra hircus (goat) "
BASE COUNT      97 a     63 c     38 g     73 t
ORIGIN

alignment_scores:
       Quality:           8.00              Length:          8
       Ratio:             1.000             Gaps:            0
Percent Similarity: 100.000       Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x X73795/rev

Align seg 1/1   to reverse of: X73795  from: 1   to: 271

    42 ArgValIleLeuPheAsnLysPro 49
       ||||||||||||||||||||||||
   224 AGAGTAATTCTTTTAACAAGCCA 201

seq_name: gb_est27:AI424418
```

```
seq_documentation_block:
LOCUS       AI424418          291 bp    mRNA         EST       28-MAR-1999
DEFINITION  te90f02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2093979 3'
            similar to gb:S37431 LAMININ RECEPTOR (HUMAN);, mRNA sequence.
ACCESSION   AI424418
VERSION     AI424418.1  GI:4270349
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 291)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Apr 21, 1998 this sequence version replaced gi:3072089.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 457   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 1.
FEATURES    Location/Qualifiers
     source  1..291
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2093979"
            /clone_lib="NCI_CGAP_Pr28"
            /sex="male"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
            with a modified polylinker; Plasmid DNA from the
            normalized library NCI_CGAP_Pr22 was prepared, and ss
            circles were made in vitro. Following HAP purification,
            this DNA was used as tracer in a subtractive hybridization
            reaction. The driver was PCR-amplified cDNAs from a pool
            of 5,000 clones made from the same library (cloneIDs
            985608-986759, 1101192-1101959, and 1217928-1220615).
            Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      68 a     76 c     80 g     67 t
ORIGIN

alignment_scores:
       Quality:           8.00              Length:          8
       Ratio:             1.000             Gaps:            0
Percent Similarity: 100.000       Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AI424418

Align seg 1/1   to: AI424418  from: 1   to: 291

    59 GluAlaGlyArgSerThrLeuLys 66
       ||||||||||||||||||||||||
   106 GAGGCTGGCGGCTCCACTCTTAAG 129

seq_name: gb_est42:AV309318

seq_documentation_block:
```

LOCUS   AV309318   304 bp   mRNA   EST   11-NOV-1999
DEFINITION   AV309318 RIKEN full-length enriched, 8 days embryo Mus musculus
             cDNA clone 5730578K03 3', mRNA sequence.
ACCESSION   AV309318
VERSION   AV309318.1   GI:6362353
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 304)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
             Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
             Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
             Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
             Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
             Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
             Shibata,Y., Shigemoto,Y., Shiraki,T., Soqabe,Y., Sugahara,Y.,
             Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N.,
             Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
             Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL   Unpublished (1999)
COMMENT   On May 18, 1998 this sequence version replaced gi:3138303.
             Contact: Yoshihide Hayashizaki
             Genome Exploration Research Group, Life Science Tsukuba Center,
             Genome Science Laboratory
             The Institute of Physical and Chemical Research (RIKEN), Genomic
             Sciences Center
             3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
             Tel: +81-298-36-9013
             Fax: +81-298-36-9098
             Email: genome-res@rtc.riken.go.jp,
             URL:http://genome.rtc.riken.go.jp/
             Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., and
             Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and
             Hayashizaki,Y.
             Transcriptional sequencing: A method for DNA sequencing using RNA
             polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,Y., Izawa,M., Kawai,J.,
             Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
             Okazaki,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
             system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
             19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
             further details.
FEATURES   Location/Qualifiers
     source   1..304
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5730578K03"
                     /clone_lib="RIKEN full-length enriched, 8 days embryo"
                     /sex="mixed"
                     /dev_stage="8 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot =100.0 Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA

                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT   93 a   82 c   61 g   68 t
ORIGIN

alignment_scores:
             Quality: 8.00   Length: 8
             Ratio: 1.000   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-252-691-7056 x AV309318/rev ..

Align seg 1/1 to reverse of: AV309318 from: 1 to: 304

        83 AspSerGluGlyLeuLeuValLeu 90
           ||||||||||||||||||||||||
        45 GACTCTGCAGGGCTTACTGGTGTTG 22

seq_name: gb_est22:AI047546

seq_documentation_block:
LOCUS   AI047546   338 bp   mRNA   EST   08-JUL-1998
DEFINITION   uh80b05.r1 Soares mouse urogenital ridge NMUR Mus musculus cDNA
             clone IMAGE:1763985 5' similar to TR:035459 O35459 ECH1P .; mRNA
             sequence.
ACCESSION   AI047546
VERSION   AI047546.1   GI:3295833
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 338)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Jan 14, 1998 this sequence version replaced gi:2754372.
             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:963509
             Possible reversed clone: similarity on wrong strand
             Seq primer: -28m13 rev2 ET from Amersham
             High quality sequence stop: 107.
FEATURES   Location/Qualifiers
     source   1..338
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1763985"
                     /clone_lib="Soares mouse urogenital ridge NMUR"
                     /sex="equal ratio of male:female"
                     /tissue_type="urogenital ridge (embryonic)"
                     /dev_stage="fetal, mixture of 11.5 and 12.5 dpc"
                     /lab_host="DH10B"
                     /note="Organ: gonad; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I: Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGCAGCGGCCGCCATTTCTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT7T3 vector.
Library went through two rounds of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."

```
BASE COUNT      74 a      86 c      100 g      78 t
ORIGIN
```

```
alignment_scores:
        Quality:   8.00          Length:      8
        Ratio:     1.000         Gaps:        0
Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
US-09-252-691-7056 x AI047546/rev     ..

Align seg 1/1  to reverse of: AI047546  from: 1  to: 338

   83 AspSerGluGlyLeuLeuValLeu 90
      ||||||||||||||||||||||||
   26 GACAGTGAAGGTCTTCTGGTACTT 3
```

seq_name: gb_est26:AI329557

```
seq_documentation_block:
LOCUS       AI329557     350 bp    mRNA              EST       28-DEC-1998
DEFINITION  b6c08ne.fl Neurospora crassa evening cDNA library Neurospora crassa
            cDNA clone b6c08ne 5', mRNA sequence.
ACCESSION   AI329557
VERSION     AI329557.1  GI:4066116
KEYWORDS    EST.
SOURCE      Neurospora crassa.
ORGANISM    Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;
            Neurospora.
REFERENCE   1  (bases 1 to 350)
AUTHORS     Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
TITLE       Two Neurospora crassa EST Databases
JOURNAL     Unpublished (1998)
COMMENT     Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            We anticipate the future release of the cDNA clones to the Fungal
            Genetics Stock Center
            Seq primer: Universal Forward Primer
            High quality sequence stop: 135.
FEATURES             Location/Qualifiers
     source          1..350
                     /organism="Neurospora crassa"
                     /strain="Strain 30-7 (bd; A)"
                     /db_xref="taxon:5141"
                     /clone="b6c08ne"
                     /clone_lib="Neurospora crassa evening cDNA library"
                     /tissue_type="tissue harvested following 22hr growth in
                     dark"
                     /note="vector: pBlueScript SK-; Site_1: XbaI; Site_2:
                     EcoRI; See: Bell-Perdersen,D., et al. PNAS 93:13096,1996.
                     5' end of cDNA cloned into XbaI site of pBluescript; 3'
                     end of cDNA cloned into EcoRI site of pBluescript"
```

```
BASE COUNT      96 a      92 c      78 g      84 t
ORIGIN
```

```
alignment_scores:
        Quality:   8.00          Length:      8
        Ratio:     1.000         Gaps:        0
Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
US-09-252-691-7056 x AI329557     ..

Align seg 1/1  to: AI329557  from: 1  to: 350

  191 PheProThrLeuArgLeuIleArg 198
      ||||||||||||||||||||||||
  209 TTCCCAACTCACGGCTTATCCGA 232
```

seq_name: gb_est38:AW047543

```
seq_documentation_block:
LOCUS       AW047543     357 bp    mRNA              EST       18-SEP-1999
DEFINITION  UI-M-BH1-ama-f-02-0-UI.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone
            UI-M-BH1-ama-f-02-0-UI 3', mRNA sequence.
ACCESSION   AW047543
VERSION     AW047543.1  GI:5908072
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 357)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3246699.
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized brain stems library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..357
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH1-ama-f-02-0-UI"
                     /clone_lib="NIH_BMAP_M_S2"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S2 library is a subtracted library derived from
                     NIH_BMAP_M_S1, which in turn is a subtracted library
                     derived from a mixture of normalized libraries from ten
                     regions of the mouse brain (cerebellum, brain stems,
                     olfactory bulbs, hypothalamus, cortex, amygdala, basal
                     ganglia, pineal gland, striatum, hippocampus). The driver
                     used for subtraction consisted of a pool of 5,000 clones
                     from the NIH_BMAP_M_S1 library and a pool of 2,000 clones
                     obtained from non-normalized and normalized mouse brain
                     spinal cord libraries.
                     TAG_LIB=NIH_BMAP_M_S2
                     TAG_TISSUE=brain-stems
                     TAG_SEQ=TCATG"
```

```
BASE COUNT      83 a      94 c      86 g      93 t      1 others
ORIGIN
```

alignment_scores:
```
        Quality:    8.00              Length:       8
        Ratio:      1.000             Gaps:         0
Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:
US-09-252-691-7056 x AW047543   ..

Align seg 1/1 to: AW047543 from: 1 to: 357

```
 83 AspSerGlyLeuLeuValLeu 90
    |||||||||||||||||||||
154 GACAGTGAAGGTCTTCTGGTACTT 177
```

seq_name: gb_est1:T52062

```
seq_documentation_block:
LOCUS       T52062         374 bp    mRNA         EST      06-FEB-1995
DEFINITION  yb29d04.r1 Stratagene fetal spleen (#937205) Homo sapiens cDNA
            clone IMAGE:72583 5' similar to contains MER6 repetitive element,
            mRNA sequence.
ACCESSION   T52062
VERSION     T52062.1  GI:653922
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 374)
AUTHORS     Hiller,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE       Generation and analysis of 280,000 human expressed sequence tags
JOURNAL     Genome Res. 6 (9), 807-828 (1996)
MEDLINE     97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 456
            High qality sequence stops: 226 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 456    Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 226.
FEATURES            Location/Qualifiers
     source         1..374
                    /organism="Homo sapiens"
                    /db_xref="GDB:494248"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:72583"
                    /clone_lib="Stratagene fetal spleen (#937205)"
                    /tissue_type="fetal spleen"
                    /dev_stage="fetal"
                    /lab_host="SOLR cells (kanamycin resistant)"
                    /note="Organ: spleen; Vector: pBluescript SK-; Site_1:
                    EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                    Oligo dT. Pooled spleens. Average insert size: 1.0 kb;
                    Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                    3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'"
BASE COUNT     89 a      88 c     109 g      85 t      3 others
ORIGIN
```

alignment_scores:

```
        Quality:    8.00              Length:       8
        Ratio:      1.000             Gaps:         0
Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:
US-09-252-691-7056 x T52062   ..

Align seg 1/1 to: T52062 from: 1 to: 374

```
 59 GluAlaGlyArgSerThrLeuLys 66
    ||||||||||||||||||||||||
218 GAGGCGGGCAGATCCACCTGAAG 241
```

us-09-252-691-7056.oligo.rst

```
ID   W94884 standard; protein; 88 AA.
AC   W94884;
DT   25-MAY-1999 (first entry)
DE   Colicin V cleaved peptide sequence.
KW   Enterocin 900; brochocin-C; bacterial growth; microorganism; inhibit;
KW   food; meat; fermentation; bacteriocin; mastitis; secretory vector; lcaC;
KW   leucocin A; lcaD: enzyme; probiotic; L. gelidum; Listeria; Lactobacillus;
KW   sulphide-producing; spoilage; sensory property; smell; taste; infection;
KW   ground beef; silage; Salmonella; poultry; intestine; therapeutic; colour;
KW   mucosal tissue; vacuum-packed; storage; colicin V.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   Disulfide_bond  76..87
FT                   /note= "disulphide bridge"
PN   WO9902555-A1.
PD   21-JAN-1999.
PF   09-JUL-1998; U14331.
PR   05-SEP-1997; US-924629.
PR   09-JUL-1997; US-052072.
PA   (UYAL-) UNIV ALBERTA.
PI   Franz C, Greer GG, Leisner JJ, McCormick JK, McMullen LM,
PI   Poon A, Stiles ME, Van Belkum MJ, Vederas JC, Worobo RJ,
PI   Worobo RW;
DR   WPI; 99-120779/10.
PT   New bacteriocins enterocin 900 and brochocin-C - useful as food
PT   preservatives and therapeutic antimicrobials
PS   Disclosure; Page 138-139; 182pp; English.
CC   The invention relates to peptides enterocin 900, brochocin-C that inhibit
CC   bacterial growth. Microorganisms expressing the above peptides are used
CC   to inhibit bacterial growth in foods (specifically meat), live animals
CC   (applied topically), food preparation areas and fermentation vessels. The
CC   bacterocins enterocin 900, brochocin-C may be used directly, specifically
CC   to inhibit bacteria that cause mastitis. Antibodies specific for the
CC   peptides are used to detect expression of bacterocins in cells. Secretory
CC   vectors (containing sequences for a promoter, a leucocin A processing
CC   peptide, a heterologous peptide, lcaC and lcaD) are used to secrete
CC   bacteriocins, enzymes or other proteins, e.g. for use in food production
CC   and as probiotics. L. gelidum, or other leucocin A-expressing bacteria,
CC   inhibit the growth of Listeria and the sulphide-producing spoilage
CC   organism Lactobacillus sake; they also improve the sensory properties
CC   (smell, taste and colour) of meat, specifically ground beef. L. gelidum
CC   can also be used in preparation of animal feeds (silage), as probiotic,
CC   to control Salmonella in poultry intestines and therapeutically against
CC   mucosal tissue infections. Treatment with bacterocins provides vacuum-
CC   packed meat products with predictable and longer storage life. The
CC   secretion vectors can express several bacteriocins for broader spectrum
CC   of activity, e.g. tailored for a particular target organism.
SQ   Sequence  88 AA;

Query Match       2.7%;  Score 6;  DB 1;  Length 88;
Best Local Similarity   100.0%;  Pred. No. 84;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     75 YAAGRL 80
          ||||||
Db     70 YAAGRL 75

Search completed: June 10, 2000, 11:27:58
Job time: 2488 sec
```